

<https://helda.helsinki.fi>

Kielitypologian menetelmät

Miestamo, Matti Markko Petteri

Suomalaisen Kirjallisuuden Seura
2020

Miestamo , M M P & Sinnemäki , K T-K 2020 , Kielitypologian menetelmät . julkaisussa M Luodonpää-Manni , M Hamunen , R Konstenius , M Miestamo , U Nikanne & K Sinnemäki (toim) , Kielentutkimuksen menetelmiä I-IV . Suomalaisen Kirjallisuuden Seuran toimituksia , Nro 1457 , Suomalaisen Kirjallisuuden Seura , Helsinki , Sivut 639-678 .

<http://hdl.handle.net/10138/323268>

cc_by_nc_nd
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Luku 14 Kielitypologian menetelmät

Matti Miestamo

 <https://orcid.org/0000-0001-9801-9030>

Kaius Sinnemäki

 <https://orcid.org/0000-0002-6972-5216>

Mikä?

Kielitypologia on tutkimusalue, joka tutkii maailman kielten monimuotoisuutta ja siihen vaikuttavia tekijöitä. Se on empiiristä kielentutkimusta, jossa pääasiallisena aineistona toimivat yksittäiskielten kielioppikuvaukset sekä rinnakkaiskorpuksien tai kyselyaineistot. Pääasiallisena teoreettisena lähtökohtana on funktionaalinen ja käyttöpohjainen lähestymistapa kieleen. Kielitypologia kehittää menetelmiä erilaisten ihmiskielten vertailuun ja tuottaa uutta perustietoa ihmiskielten luonteesta ja kielten variaatiosta.

Katso myös:

Luku 12 Kielikontaktien tutkimus

Luku 13 Kielten dokumentointi ja kieliopin kuvaus

Luku 15 Historiallinen ja vertaileva kielentutkimus

Luku 17 Kieliopin tutkimus ja kielioppiteoriat

1. Johdanto

Tässä artikkelissa pureudumme kielitypologian teoriataustaan ja metodologisiin ratkaisuihin.¹ Kielitypologia vertailee maailman kieliä usein hyvinkin laajojen otosten pohjalta ja pyrkii tätä kautta ymmärtämään kielten ominaisuuksia ja kielen olemusta. Se on siis teoreettista kielitiedettä, jonka empiirinen perusta on maailmanlaajuisessa kielten vertailussa. Keskeisiä kysymyksiä kielitypologille ovat: Mikä on maailman kielille yhteistä – universaalia – ja miten kielet eroavat toisistaan; siis mitkä ovat kieltenvälisen variaation rajat? Mikä on yleistä ja mikä harvinaista maailman kielissä? Millaisia yhteyksiä eli korrelaatioita eri kielenpiirteiden välillä voidaan havaita (kielenpiirteen määrittelystä ks. luku 3.4.)? Miten eri rakennetyypit jakautuvat maantieteellisesti? Kuten millä tahansa muulla (kieli)tieteen alalla näihin kaikkiin kysymyksiin lisätään myös soveltuva miksi-kysymys, eli pyritään selittämään ja ymmärtämään tehtyjä kieltenvälisiä havaintoja ja yleistyksiä. Nimitys *typologia*, joka juontaa juurensa tutkimusalan historiaan, korostaa ehkä liiaksikin luokittelun merkitystä – luokittelu on yksi typologisen tutkimuksen vaihe ja osatavoite matkalla ymmärrykseen ihmiskielestä.

Varsinainen kielitypologinen tutkimus on nimenomaan laajoihin kieliotoksiin pohjautuvaa tutkimusta. Väljemmin ymmärrettynä typologista tai typologisesti suuntautunutta tutkimusta voidaan kuitenkin tehdä myös suppeamman kielten vertailun pohjalta tai vaikkapa yksittäistä kieliryhmää tai kieltä typologisesta näkökulmasta tarkastellen. Tässä artikkelissa keskitymme ensisijaisesti laajaan vertailuun perustuvan typologisen tutkimuksen metodeihin. Suomeksi kielitypologian metodologiasta, erityisesti kielten vertailtavuuden problematiikan kannalta, on aiemmin kirjoittanut Miestamo (2013); käsillä olevassa artikkelissa laajennamme näkökulmaa ja omistamme enemmän tilaa tutkimusprosessin muille vaiheille, mm. aineiston keruulle kielten laajassa vertailussa.

Periaatteessa mitä tahansa kielen ilmiötä voidaan lähestyä laajasta vertailevasta näkökulmasta. Käytännössä eniten typologista tutkimusta on tehty kieliopin keskeisistä osa-alueista, kuten kielten morfologisista ominaisuuksista, sanajärjestyksestä, lauseen osallistujien merkinnästä

ja tempus-, aspekti- ja moduskategorioista. Tällaisia ilmiöitä on metodologisesti helpompi lähestyä, koska niistä yleensä löytyy tietoa hyvinkin suuresta määrästä kieliä. Kielten peruskuvaukset käsittelevät useimmiten ainakin tällaisia keskeisiä kieliopin osa-alueita. On siis luonnollista, että kielitypologinen tutkimus on lähtenyt liikkeelle näistä ilmiöistä. Foneettinen ja fonologinen tutkimus on perinteisesti aina ollut kieltenvälisesti orientoitunutta, vaikka sitä onkin harjoitettu jossain määrin erillään varsinaisesta kielitypologisesta tiedeyhteisöstä. Nytemmin fonologinen typologia on enenevässä määrin integroitumassa typologian muuhun kenttään. Myös leksikaalinen typologia eli sanamerkitysten ja leksikon ominaisuuksien laaja kieltenvälinen vertailu on viime aikoina saanut kasvavaa huomiota typologiassa. Sellaisia kielen ilmiöitä, joiden kieltenvälinen vertailu vaatii typologilta yksittäisten kielten aineiston syvällistä analyysia, on vaikeampi tarkastella laajasta typologisesta näkökulmasta. Niinpä esimerkiksi erilaisten diskurssi-ilmiöiden typologinen tutkimus on vielä lapsenkengissään. Hyvän kuvan kielitypologiassa perinteisesti käsitellyistä aihepiireistä antaa tutustuminen teoksen *The World Atlas of Language Structures* verkkoversioon osoitteessa <https://wals.info> (Dryer & Haspelmath 2013).

Yleinen kielitiede pyrkii tekemään kaikkia kieliä koskevia yleistyksiä. Laajamittaista kielten vertailua harjoittava kielitypologia on näin yleisen kielitieteen kentän keskiössä. Kaikkia kieliä koskevia yleistyksiä ei voida tehdä yhden tai vain muutaman kielen aineiston pohjalta. Yleisesti maailman kieliin soveltuvia teorioita ja metodeja ei voida kehittää ottamatta huomioon kieltenvälistä variaatiota maailmanlaajuisessa perspektiivissä. Niillekin tutkimusasetelmille, joissa empiirinen perusta ei käsitä laajaa kielten vertailua, on tärkeää ottaa huomioon kielitypologiassa tehdyt havainnot ja eri kielistä löytyvät ilmiöt. Kielitypologia ruokkii näin teorianmuodostusta kielitieteen eri aloilla. Laajemmassa tieteidenvälisessä kontekstissa voidaan todeta, että pyrkiessään kaikkia kieliä koskeviin yleistyksiin yleinen kielitiede – ja erityisesti kielitypologia – on tärkeä osa sitä monimutkaista palapeliä, jota tieteentekijät kokoavat ymmärtääkseen ihmismieltä ja ihmisen toimintaa.

Kielitypologian ja yksittäiskielten tutkimuksen välillä vallitsee riippuvuussuhde ja hedelmällinen vuorovaikutus. Toisaalta kielitypologit

ovat riippuvaisia yksittäiskielten tutkijoiden tuloksista voidakseen ylipäätään vertailla kieltä. Toisaalta yksittäiskielten tutkijat hyötyvät typologisesta tiedosta, joka auttaa ymmärtämään yksittäiskielen ilmiöitä laajemmassa kontekstissa ja antaa näin eväitä analyysien hiomiselle ja aineiston uudelleentulkinnalle. Mitä tahansa kielen ilmiötä yksittäisessä kielessä tai kieliryhmässä tutkiessaan kannattaa aina tutustua vastaaviin ilmiöihin muissa kielissä, ja mikäli ilmiöstä on olemassa typologista tutkimusta, se tarjoaa helpoimman pääsyn tällaiseen kieltenväliseen tietoon. Pyrimme tässä artikkelissa valottamaan typologian teoreettista ja metodologista taustaa, ei pelkästään niille, joiden oma tutkimus on typologista luonteeltaan, vaan myös niille, jotka käyttävät typologian tutkimustuloksia oman yksittäisiä kieliä koskevan tutkimuksensa taustalla. Tutustuminen typologian teoreettis-metodologisiin taustoihin auttaa ymmärtämään typologisen tiedon luonnetta syvällisemmin.

Artikkeli etenee seuraavasti. Luvussa 2 luomme lyhyen katsauksen kielitypologian historiaan ja historialliseen taustaan sekä muutamisiin keskeisiin teoreettisiin kysymyksiin. Luvussa 3 tarkastelemme typologisia tutkimusprosesseja: vertailuperustaa, otoskielten valintaa, luokittelua, jakaumien tarkastelua, kielenpiirteiden välisiä korrelaatioita ja tulosten selittämistä. Luvussa 4 pohdimme lyhyesti typologian paikkaa kielitieteen kentällä suhteessa lähialoihin sekä teemme lyhyen yhteenvedon artikkelin teemoista.

2. Kielitypologian historiaa ja teoreettista taustaa

2.1. Katsaus kielitypologian historiaan

Kielten systemaattinen ja laajamittainen vertailu on lähestymistapana suhteellisen nuori. Se alkoi varsinaisesti Joseph Greenbergin tutkimuksista 1960-luvulla. Kaikkien kielten yhteyttä etsivä universalistinen ajattelu on sen sijaan paljon vanhempaa perua eurooppalaisessa

kontekstissa. Se juontaa juurensa jo antiikin ajattelijoihin, ja omat universaali-kielioppi-mallinsa ovat myöhemmin esittäneet mm. modistien nimellä tunnettu koulukunta myöhäiskeskiajalla sekä Port Royal-koulukunta, joka vaikutti 1600-luvulla Ranskassa (ks. Itkonen 1991). Yhteistä näille historiallisille suuntauksille on, että universaaleista päämääristä huolimatta tarkastelu on rajoittunut lähinnä kulloiseenkin valtakieleen. Tässä mielessä ne vertautuvat suoraan chomskylaiseen generatiiviseen kielioppiin, joka nykysuuntauksista ehkä selvimmin perustaa olettamuksensa kielen universaaleista ominaisuuksista yhden (tai korkeintaan muutaman) kielen analyysille. Kielitypologian lähtökohta on täysin vastakkainen: tutkittaessa maailman kielten erilaisuutta ja samuutta on empiirisen aineiston katettava maailman kielten kirjo mahdollisimman laaja-alaisesti ja tasaisesti.

Kielten vertailusta alettiin kiinnostua enenevässä määrin 1800-luvulle tultaessa. Keskeisimpiä huomion kohteita oli kielten morfologinen rakenne. Klassinen morfologinen typologia, jonka näkyvimpiä edustajia olivat mm. August (1767–1845) ja Wilhelm (1772–1829) von Schlegel sekä Wilhelm von Humboldt (1767–1835), jakoi kielet kolmeen perustyyppiin: isoloiviin kieliin, joissa ei ole sidonnaista morfologiaa lainkaan eli kukin sana koostuu yhdestä morfeemista, sekä agglutinoiviin ja fusionsaaliisiin kieliin, joissa molemmissa sanat sisältävät affikseja, mutta edellisissä vartaloiden ja affiksien muodoissa ei ole vaihtelua, kun taas jälkimmäisissä esiintyy erilaisia morfofonologisia ilmiöitä. Myöhemmin näihin kolmeen perustyyppiin lisättiin vielä neljäs eli polysyn-teettiset kielet, joissa affiksien määrä sanassa on huomattavan suuri ja yksi sana voi äärimmillään vastata pitkääkin lausetta kolmen muun tyyppin kielissä. Huomattakoon, että kun puhuttiin kielityypeistä eikä rakennetyypeistä, ajatuksena oli, että kielen morfologinen tyyppi luonnehtisi kieltä yleisellä tasolla. Tuon ajan typologiseen ajatteluun liittyi myös etnopsykologinen ote, jossa kielen tyyppin ajateltiin heijastelevan puhujayhteisön kulttuurista kehitystasetta ja ajattelutapaa. 1800-luvun lopulla ja 1900-luvun alkupuoliskolla kiinnostus typologiseen vertailuun oli vähäisempää; 1800-luvulla dominoi historiallis-vertaileva kielitiede, ja 1900-luvulla siirryttiin vähitellen strukturalismiin. Typologista ajattelua veivät kuitenkin strukturalismin rinnalla eteenpäin esimerkiksi

Edward Sapir (1884–1939), joka kehitti mm. morfologisen typologian luokitteluita analyttisempaan suuntaan tunnistamalla erillisiä parametrejä aiemmin yhtenäisempinä esitettyjen morfologisten tyyppien takana (ks. Sapir 1921), sekä Roman Jakobson (1896–1982) ja muutamat muut Prahan koulukunnan edustajat.

Kiinnostus kielten vertailuun ja kielten universaaleihin ominaisuuksiin heräsi toden teolla 1960-luvulla, kun Joseph Greenberg (1915–2001) toi laajan systemaattisen kielten vertailun metodit kielitieteeseen. Eräänlaisena lähtölaukauksena laajempaan tietoisuuteen voidaan pitää Dobbs Ferryssä vuonna 1961 järjestettyä konferenssia *Universals of Language*, johon Greenberg oli kutsunut useita tunnettuja kielitieteilijöitä ja lähialojen edustajia. Tuon konferenssin pohjalta toimitetussa kirjassa ilmestyi Greenbergin (1963) artikkeli ”Some universals of grammar with particular reference to the order of meaningful elements”, joka loi pohjan systemaattiseen kielen vertailuun perustuvalle typologiselle metodologialle ja greenbergiläiselle universaalikäsitelykselle. Vuosina 1967–1976 Greenberg johti Stanfordissa *Universals of Human Language* -projektia, jonka tulokset on vedetty yhteen neliosaisessa kirjasarjassa *Universals of Human Language* (Greenberg 1978). Näiden tapahtumien myötä moderni kielitypologia oli saanut alkunsa (Greenbergistä ks. tarkemmin Miestamo 2010).

Myöhemmistä kielitypologian kehityksen virstanpylväistä mainittakoon seuraavat. 1990-luvulla EU:n rahoittaman EURO-TYP-projektin tuloksena syntyi kokoelma kirjoja, joissa kartoitettiin laajasti Euroopan kielten ominaisuuksia kieliopin eri osa-alueilla. Euroopan kieliä tarkasteltiin maailmanlaajuisessa typologisessa perspektiivissä, joten samalla tehtiin laajempaakin typologista tutkimusta. Vuonna 1994 perustettu Association for Linguistic Typology aloitti säännöllisen konferenssisarjansa vuonna 1995 ja alkoi julkaista *Linguistic Typology* -lehteä vuonna 1997. Leipzigin evoluutioantropologian Max Planck -instituutissa vuonna 1997 aloittanut kielitieteen laitos muodostui merkittäväksi kielitypologisen tutkimuksen keskuksesi – Max Planckin politiikan mukaisesti laitos lakkautettiin vuonna 2015 sen johtajan Bernard Comrien jäädessä eläkkeelle. Vuonna 2005 ilmestyi edellä mainittu, noin 40 kirjoittajan yhteistyönä maailman kielten ominaisuuksia kartoittava *The World Atlas*

of *Language Structures* (WALS) (Haspelmath, Dryer, Gil & Comrie 2005). WALS on näkyvä esimerkki siitä viimeaikaisesta kehityksestä, jossa kielten ja kielenpiirteiden maantieteelliseen levinneisyyteen liittyvät kysymykset ovat enenevässä määrin nousseet kielitypologian keskiöön. Moderni kielitypologia syntyi 1960-luvulla vastaamaan kysymykseen kieltenvälisen variaation rajoista ja kielten universaaleista eli siitä, millaiset kielet ovat mahdollisia ja millaiset eivät ja miten tätä kieltenvälistä variaatiota voidaan selittää (etupäässä) funktionaalisesti. Vuosituhannen vaihteen jälkeen on yleistynyt Johanna Nicholsin (1992) klassikkoteoksesta lähtölaukauksensa saanut populaatiotieteellinen näkökulma, jossa rakennetyyppien levinneisyyttä tarkastellaan populaatioina. Bickelin (2007) sanoin kysymys ”mikä on kielessä mahdollista?” on saanut seurakseen kysymykset ”mitä, missä ja miksi?”. Siinä missä ennen otannassa pyrittiin eliminoimaan kielisukulaisuuteen ja kielikontakteihin perustuvat yhtäläisyydet, jotta päästäisiin käsiksi kielen perimmäisiin ominaisuuksiin, nykyisin areaalinen ja genealoginen näkökulma ovat itessään usein mukana typologisessa tutkimuksessa, ja tyyppien yleisyyttä ja levinneisyyttä halutaan ymmärtää niiden taustalla olevien historiallisten prosessien kautta (ks. tarkemmin Bickel 2007; [ks. Kielikontaktien tutkimus tässä kirjassa](#)). Palaamme näihin eri suuntauksiin luvussa 3. Kielitypologian historiasta voi lukea lisää esimerkiksi Ramatin (2010) ja Graffin (2010) katsausartikkeleista.

Suomalaisen kielitieteen kontekstissa voidaan typologisesti orientoituneen tutkimuksen varhaisia juuria nähdä jo 1800-luvulla Anders Sjögrenin ja Matthias A. Castrénin itään suuntautuneissa tutkimusmatkoissa ([ks. myös Antropologinen lingvistiikka tk.](#)). Ne aloittivat kansainvälisesti urauurtavan etnolingvistisen tradition, jossa tarkasteltiin ja vertailtiin eurooppalaisista kielistä typologisesti monella tavalla poikkeavia kieliä ja kulttuureita. Tuloksena syntyi valtavasti typologisesti kiinnostavaa aineistoa pohjoisen Euraasian kielistä. Vertailu oli kuitenkin päämääriltään ensisijaisesti historiallis-vertailevaa, ei typologista. Fennougriistiikassa ja fennistiikassa diakroninen näkökulma dominoi aina 1960-luvulle asti. Modernin typologian aikakaudella typologinen näkökulma alkoi rantautua Suomeen. Mikko Korhosen tuotannossa kielitypologinen näkökulma oli vahvasti läsnä (ks. Salminen 1996).

Yllämainittuun EUROTYP-hankkeeseen osallistui suomalaisista kielitieteilijöistä mm. Marja Leinonen, Jouko Lindstedt, Hannu Tommola ja Maria Vilkuna. Turun yliopiston yleisen kielitieteen oppiaineessa Esa Itkonen ja Anneli Pajunen toivat typologista näkökulmaa vahvasti esille (ks. Pajunen 1998; 2000; Itkonen 1997). Pirkko Suihkonen pohti väitöskirjassaan (1990) korpustutkimuksen mahdollisuuksia kielitypologiassa. Ensimmäisiä varsinaisia laajoihin otoksiin perustuvia typologisia tutkimuksia olivat Seppo Kittilän transitiivisuutta (2002) käsittelevä ja Matti Miestamon negaatiota (2000; 2003) käsittelevä tutkimus. Kvantitatiivista näkökulmaa suomalaiseen kielitypologiaan on tuonut erityisesti Kaius Sinnemäki (2008; 2011). Annu Marttilan (2010) leksikaalista ikonisuutta lintujen nimissä tarkastelevaa laajaa vertailevaa tutkimusta voidaan pitää leksikaalisen typologian alaan kuuluvana. Myös Ksenia Shagalin (2019) tuore tutkimus partisiippien typologiasta on herättänyt kansainvälistä huomiota.

2.2. Teoreettista taustaa

Ylempänä typologiaa on luonnehdittu maailmanlaajuiseksi vertailevaksi kielitieteeksi ja teoreettiseksi kielitieteeksi, jonka empiirinen perusta on maailmanlaajuisessa kielten vertailussa. Nämä luonnehdistinnat sinällään eivät vielä kerro paljoakaan siitä, millaisia teoreettisia taustaoletuksia typologisella tutkimuksella on. Toki empiirinen metodologinen lähestymistapa asemoi typologian rationalismi-empirismijaottelussa empirismin puolelle. Periaatteessa laajaan vertailuun perustuvaa typologista tutkimusta voidaan tehdä hyvinkin erilaisista teoreettisista lähtökohdistta. Tavallisimmin kuitenkin typologista tutkimusta harjoitetaan funktionalistisessa viitekehyksessä. Myöskään formaaleja metakieliä ei yleensä käytetä, vaan analyysit esitetään mahdollisimman laajasti ymmärretyin ja jaetun käsitteistön puitteissa. Typologinen vertailu on luonteeltaan synkronista, mutta diakroninen näkökulma on vahvasti läsnä typologisessa tutkimuksessa, etenkin typologisten havaintojen selittämisessä. Keskeinen teoreettinen kysymys on myös yksittäiskielten ja universaalien suhde ja kielellisten kategorioiden luonne kielikohtaisina vs. universaaleina

sekä edelleen se, miten erilaiset kategoriäkäsitteet vaikuttavat kielen vertailtavuuteen konkreettisen metodin tasolla. Tässä luvussa luomme lyhyen katsauksen kielitypologian keskeiseen teoreettiseen taustaan.

Kun eri lähestymistapoja kieleen ja kielioppiin luonnehditaan teoreettisen lähestymistavan mukaan joko formalistisiksi tai funktionalistisiksi, typologia identifioituu yleensä selvästi jälkimmäiseen ryhmään ([ks. Johdannon luku 3 tk.](#)). Kuten yllä todettiin, laajaa vertailua voidaan toki tehdä monenlaisissa teoriakehyksissä, ja esimerkiksi generatiivinen kielitiede on viime aikoina alkanut kiinnittää enemmän huomiota kieltenväliseen evidenssiin. Laajat vertailevat tutkimukset tällaisista formalistisista lähtökohdista ovat kuitenkin harvinaisia. Yleensä typologioiden keskeisenä teoreettisena lähtökohtana on, että kielen funktio, kielenkäyttö, vaikuttaa kielen rakenteeseen, ja näin selityksiä kieltenvälisille yleistyksille etsitään ensisijaisesti funktion tasolta eli semantiikasta, pragmatiikasta ja kognitiivisista prosesseista. Jo Greenbergillä (1966) oli vahvasti käyttöpohjainen näkemys kielestä, mistä selkeimpänä esimerkkinä on se, että hän nosti erityisesti ilmiöiden frekvenssin tärkeäksi selittäväksi tekijäksi. Typologian selityspäätteisiin palaamme luvussa 3.7.

Kielten laajahkoa vertailua harjoittavat myös historiallis-vertailevan kielitieteen edustajat. Typologinen vertailu eroaa tästä tavoitteidensa osalta: typologiassa vertailun päämäärä ei ole päästä kiinni kielten historialliseen kehitykseen ja sukulaisuussuhteisiin vaan kuvata kieltenvälistä variaatiota ja ymmärtää ihmiskielen perusperiaatteita laajan vertailun pohjalta. Typologia on siis ensisijaisesti synkronista, siinä missä historiallis-vertailevan kielitieteen tavoitteet ovat lähtökohtaisesti diakroniset. Diakronialla on kuitenkin keskeinen rooli typologiassa. Jo Greenberg (1978) puhui dynaamisesta typologiasta ja tarkasteli mm. sitä, millaiset muutokset rakennetyypistä toiseen ovat mahdollisia, ja tällaiset kysymyksenasettelut ovat tulleet entistä tärkeämmiksi makrotason synkronisten jakaumien selittäjinä. Havaittujen rakennetyyppien jakaumien selittämisessä on oleellista katsoa niiden historiallisen kehityksen yksityiskohtia ja tyypillisiä kieliopillistumiskehityksiä. Vaikka itse vertailu onkin yleensä synkronista, typologisissa tutkimuksissa on usein vahva diakroninen komponentti mukana. Kielen funktionaaliset aspektit vaikuttavat kielen muotoon ja rakenteeseen historiallisten prosessien

kautta, ja typologiien selitysmalleissa yhdistyvät näin historialliset ja funktionaaliset selitykset.

Typologiassa, kuten myöskään deskriptiivisessä kielitieteessä ([ks. Kielten dokumentointi ja kieliopin kuvaus tk.](#)), ei yleensä käytetä teoriaspesifejä, formalisoituja metakieliä. Sen sijaan analyysit esitetään käyttäen sellaista terminologiaa, joka on kielitieteilijöiden keskuudessa mahdollisimman laajalti jaettua ja ymmärrettyä. Dixon (1997; 2009) ja Dryer (2006) ovat esittäneet, että tällaista yhteistä teoriapohjaa voidaan kutsua *peruskielioppiteoriaksi* (engl. *Basic Linguistic Theory*, BLT). Peruskielioppi-teorialla on juurensa perinteisessä kieliopissa, jonka päälle se on eri aikakausien kielitieteellisen tutkimuksen tuloksia hyödyntäen rakentanut modernin, eri kielten ilmiöiden analyysiin ja kuvaukseen laajalti soveltuvan käsitteistön ja terminologian. Peruskielioppiteoria on yhteinen, kumulatiivinen teoriapohja, jonka osaksi pääsee kulloinkin vallalla olevista teorioista käsitteitä, jotka osoittautuvat kestäviksi ja hyödyllisiksi; jotkin käsitteet taas elävät vain lyhyen aikaa eivätkä päädy osaksi peruskielioppi-teoriaa. Nykyisin peruskielioppiteoria mainitaan usein eksplisiittisesti teoriataustaksi typologisissa tutkimuksissa ja deskriptiivisissä kieliopeissa.

Keskustelussa typologisen vertailun teoreettisista lähtökohdista on viime aikoina noussut keskeisenä teemana esiin pohdinta kielellisten kategorioiden luonteesta: ovatko kielelliset kategoriat kielille yhteisiä vai kielikohtaisia. **Universalistinen** kategoriakäsitys olettaa kaikille kielille yhteisten kategorioiden olemassaolon, kun taas **partikularistinen** käsitys pitää kielellisiä kategorioita kielikohtaisina. Edellisen mukaan siis sellaiset kategoriat kuin vaikkapa adjektiivi, subjekti tai genetiivi olisivat kielille yhteisiä ja kunkin yksittäisen kielen kategoriat olisivat näiden yhteisten kategorioiden kielikohtaisia reaalistumia. Jälkimmäinen käsitys taas olettaa, että kategorioita ei kieltenvälisessä merkityksessä ole olemassa, vaan ne ovat aina kielikohtaisia. Universalistinen kategoriakäsitys on vallalla nykykielitieteessä varsinkin erilaisten formalististen kielioppimallien kannattajien parissa. Niistä erityisesti mainittakoon chomskylainen generativismi, jossa kategoriat ovat synnynnäisen universaali-kieliopin määäämiä. Myös esimerkiksi perinteinen, antiikin kielitieteestä juurensa juontava ja 1800-luvulle asti yleinen tapa kuvata eri kieliä latinan mallin mukaisesti voidaan nähdä universalistisena

lähestymistapana. Partikularistinen kategoriakäsitys on ollut erityisen näkyvä 1900-luvun alkupuolen amerikkalaisessa antropologisessa kielitieteessä ja strukturalismissa, jossa tutkijat saivat tarkasteltavakseen valtavan kirjon eurooppalaisista kielistä radikaalistikin poikkeavia kieliä; reaktio tähän kielten moninaisuuteen oli Martin Joosin (1957, 96) sanoin näkemys, että ”kielet voivat erota toisistaan rajattomasti ja ennustamattomasti”.² Chomskylaisen generativismin nousu käänsi valtavirta-ajattelua vahvasti universalistiseen suuntaan, mutta viime vuosikymmeninä funktionalistisessa kielitieteessä ja etenkin typologiasa on otettu jälleen eksplisiittisesti partikularistisia kantoja (ks. esim. Lazard 1992; Dryer 1997; Haspelmath 2007). Funktionaalis-typologinen partikularistinen näkemys ei kuitenkaan ole aiempien amerikkalaisten strukturalistien tavoin, että kielten välinen variaatio olisi satunnaista. Kategoriat ovat kielikohtaisia, mutta niiden välillä on yhtäläisyyksiä ja samankaltaisuutta, jonka taustalla on kielille yhteinen funktionaalinen taso – merkitys, käyttö ja kommunikaatiotarpeet. Tutkijan käsitys kategorioiden luonteesta vaikuttaa suoraan vertailun metodologisiin perusteisiin, joihin palataan seuraavassa luvussa.

3. Typologinen tutkimusprosessi

Tässä luvussa tarkastelemme lähemmin typologisen tutkimusprosessin eri vaihteita. Funktionalistis-typologinen tutkimusprosessi voidaan tyypillisesti tiivistää alla olevaan seitsemään eri vaiheeseen (soveltaen Stassen 1985, 1–23; Miestamo 2013). Tässä esitetyjä vaihteita edeltää luonnollisesti tutkimuksen tavoitteiden ja tutkimuskysymysten määrittely (suluissa luku, jossa käsittelemme kyseistä tutkimusprosessin vaihetta).

1) Typologisen tutkimusprosessin vaiheet

I. Aineiston koonti

1. Määritellään tutkimuksen kohde (3.1.)

2. Valitaan kieliotos (3.2.)

3. Hankitaan tutkimusmateriaali (3.3.)
- II. Aineiston analyysi ja yleistysten tekeminen
 1. Analysoidaan ja luokitellaan materiaali (3.4.)
 2. Kuvaillaan rakennetyyppien yleisyys ja maantieteellinen jakauma (3.5.)
 3. Selvitetään tilastolliset korrelaatiot (3.6.)
- III. Tulosten selittäminen (3.7.)

Tämä representaatio on toki yksinkertaistus käytännön työstä. Tutkimus ei useinkaan etene näin suoraviivaisesti, vaan eri vaiheita voidaan työstää samanaikaisesti tai aiempiin vaiheisiin voidaan joutua palaamaan useamman kerran. Analyysi- ja luokitteluvaiheessa saatetaan palata materiaalin keräämiseen, jos alkuperäinen lähde ei ollut riittävä esiin nousseiden kysymysten ratkaisemiseksi. Jos analyysivaiheessa lähde-materiaali osoittautuu muutoinkin riittämättömäksi, typologi voi joutua vaihtamaan otoskielen ja täten palaamaan otantavaiheeseen. Joskus analyysivaiheessa täytyy myös tarkentaa itse tutkimuskohteen määrittelyä, jos se ei ole osoittautunut riittävän tarkaksi. Seuraavassa kuvaamme tutkimusprosessia vaihe vaiheelta.

3.1. Tutkimuskohteen määrittely

Typologinen tutkimus lähtee liikkeelle tutkimuskohteen määrittelystä. Oleellinen kysymys on, miten tunnistaa tutkimuksen kohteena oleva ilmiö hyvinkin erilaisissa kielissä, jotta vertailu kohdistuu samaan ilmiöön vertailtavissa kielissä. Tätä ongelmaa kutsutaan typologisessa kirjallisuudessa *kieltenvälisen identifikaation ongelmaksi* (ks. esim. Stassen 2010). Edellisessä luvussa totesimme, että typologisen tutkimuksen yleinen teoreettinen taustaoletus on, että kielelliset kategoriat ovat kielikohtaisia, siis kategoriakäsitys on partikularistinen. Kun kielille yhteisiä kategorioita ei ole olemassa, ei vertailtavan tutkimuskohteen määrittelyä voida perustaa sellaisille.

Jo varhain todettiin, että kielellisen vertailun perustaminen puhtaasti muodollisiin kategorioihin on ongelmallista (Greenberg 1963; Stassen

1985). Mikäli halutaan vertailla eri kielten adjektiiveja, ei niitä voida tunnistaa sellaisten yhdestä tai muutamasta kielestä peräisin olevien muo-
dolisten kriteerien avulla kuin että niillä on morfologisesti ilmaistut ver-
tailuasteet positiivi (*hyvä*), komparatiivi (*parempi*) ja superlatiivi (*paras*),
että ne tarvitsevat predikaatteina toimiessaan *olla*-verbin (esim. *Erkki
on pitkä*) ja että ne voivat esiintyä intensifikaattorien, kuten *erittäin* ja
sangen, kanssa. Ominaisuussanojen ilmaisussa on paljon kieltenvälistä
variaatiota, ja tällainen muotoon perustuva määritelmä jättää ulko-
puolelleen esimerkiksi ne lukuisat kielet, joissa ominaisuussanojen ver-
tailu toteutetaan muuten kuin morfologisesti tai joissa ominaisuussanat
eivät ole nomineja vaan toimivat syntaktisesti kuten verbit. Jo varhain
siis havaittiin, että kieltenvälisen vertailun on tavallisesti perustuttava
muodon sijasta funktioon (ks. Stassen 2010; Miestamo 2013).

Tätä metodologista käytäntöä on viime aikoina kirkastanut Haspelmathin
(2010) eksplikoima kuvauskategorioiden (engl. *descriptive categories*) ja
vertailukäsitteiden (engl. *comparative concepts*) välinen eronteko. Kuvaus-
kategorioiden kohteena ovat yksittäisten kielten kielikohtaiset kategoriat,
joiden semanttiset ja morfosyntaktiset ominaisuudet ovat kullekin
kielelle uniikkeja. Nämä ovat osa kunkin kielen systeemiä, ja ne ovat
psykologisesti ja sosiaalisesti tosia siinä mielessä, että ne ovat olemassa
kunkin kielen puhujayhteisön jakamina konventioina. Kielentutkijan
esittämien kuvauskategorioiden psykologinen ja sosiaalinen totuus puo-
lestaan riippuu siitä, miten hyvin tutkija on onnistunut tiivistämään to-
dellisuuden esittämiinsä kategorioihin. Vertailukäsitteet sen sijaan ovat
tutkijan metodologisia työkaluja, joiden tavoitteena on toimia kielen-
välisten yleistysten välineenä, eikä niillä ole suoraa relevanssia minkään
yksittäisen kielen kuvaukselle. Epäsuorasti ne toki voivat olla hyvinkin
relevantteja – kuten alussa totesimme, typologisesta tiedosta on paljon
hyötyä kielenkuvauksen prosessissa (ks. myös luku 4). Siinä missä tut-
kijan esittämät kuvauskategoriat voivat olla oikeita tai vääriä suhteessa
kunkin kielen todellisiin kielikohtaisiin kategorioihin, vertailukäsitteet
eivät ole samanlaisessa suhteessa todellisuuden kanssa. Metodologisina
työkaluina niitä arvioidaan ensisijaisesti sen mukaan, miten hyvin ne
soveltuvat kulloisenkin tutkimuskysymyksen selvittelyyn, esimerkiksi
miten selkeästi ja luotettavasti vertailukäsite rajaa tutkittavan ilmiön

hyvinkin erilaisista otoskielistä ja missä määrin sen määritelmään perustuva tutkimus tuottaa teoreettisesti kiinnostavia tuloksia.

Kielten kategoriat ovat kielikohtaisia, mutta niiden välillä voidaan havaita paljonkin samankaltaisuutta. Funktionalistisessa teoriakehyksessä kielten kategorioiden samankaltaisuuden nähdään perustuvan kielten keskenään jakamiin funktionaalsiin ominaisuuksiin. Kieltenvälinen vertailu pohjaa näin ollen ensisijaisesti universaalien tason semanttisiin ja pragmaattisiin funktioihin, joita voidaan lisäksi tarkentaa muotoon liittyvillä kriteereillä. Haspelmathin (2010, 670) mukaan vertailukäsitteiden määrittelyssä voidaan käyttää kolmenlaisia käsitetyyppejä:

- konseptuaalis-semanttiset käsitteet (esim. *ominaisuus* alla olevan laatikon esimerkkimääritelmässä)
- yleiset muodolliset käsitteet (esim. *edeltävä* määriteltäessä esim. prepositiota)
- muut vertailukäsitteet (esim. *lekseemi* alla olevan laatikon esimerkkimääritelmässä).

Laatikon esimerkki havainnollistaa vertailukäsitteen määrittelyä.

Esimerkki:

Haspelmath (2010, 670) määrittelee vertailukäsitteen *adjektiivi* seuraavasti:

Adjektiivi on lekseemi, joka merkitsee deskriptiivistä ominaisuutta ja jota voidaan käyttää rajoittamaan substantiivin tarkoitetta.

Määritelmässä esiintyvät vertailukäsitteet *lekseemi* ja *substantiivi* sekä konseptuaalis-semanttiset käsitteet *deskriptiivinen ominaisuus* ja *rajoittaa tarkoitetta*. Tämä määritelmä rajaa ulkopuolelle esimerkiksi sanat *kauneus* ja *koko*, jotka ovat kyllä lekseemejä ja viittaavat ominaisuuteen mutta joita ei käytetä substantiivin määritteenä rajaamaan sen tarkoitetta. Määritelmä rajaa ulkopuolelle myös konstruktion *joka oli kaunis*, joka kyllä viittaa ominaisuuteen ja rajaa substantiivin tarkoitetta mutta joka ei ole lekseemi vaan relatiivilause.

Kieltenvälisen identifi kaation on gelma ja kielten vertailtavuus ovat keskeisimpiä metodologisia kysymyksiä kielitypologiassa. Emme kuitenkaan kirjoita niistä tässä tämän pidemmin, koska Miestamon (2013) suomenkielinen esitys aiheesta on avannut problematiikkaa konkreettisesti ja runsain esimerkein.

3.2. Otanta

Kun tutkimuskohde on määritelty, seuraava vaihe tutkimusprosessissa on valita kielet, joista tutkimusmateriaali kerätään ja analysoidaan. Ei ole kuitenkaan itsestään selvää, mitkä maailman noin 7 000 kielestä pitäisi valita tutkimukseen. Kielten systemaattista valintaa ja siihen liittyviä periaatteita kutsutaan *otannaksi*, ja se on laajimmin käsitelty aihe typologisessa metodologiassa (ks. esim. Cysouw 2005, 554–555). Joskus eksploraatiivisessa eli aiemmin heikosti tunnetun tai tuntemattoman ilmiön kartoittavassa tutkimuksessa käytetään ns. mukavuusotantaa, eli kieliä otetaan mukaan tutkimukseen ilman systemaattisia otantaperiaatteita, mutta tämä ei ole nähdäksemme suositeltavaa edes tutkimuksen pilottivaiheessa. Otannassa kielten valintaa siis rajoitetaan etukäteen erilaisin periaattein, jotta saadun otoksen avulla voitaisiin tehdä yleistyksiä.

Typologiassa käytetyt otantamenetelmät voidaan jakaa karkeasti kahteen luokkaan niiden tarkoituksen mukaan: variaatio-otantaan ja todennäköisyysotantaan. Variaatio-otantaa käytetään, kun pyritään kartoittamaan jonkin kielenpiirteen kieltenvälinen variaatio. Tällöin on tavoitteena saada mahdollisimman kattava käsitys vertailukäsitteen avulla määritellyn piirteen monimuotoisuudesta ja myös sen kaikkein harvinaisimmista ilmenemistyypeistä. Variaatio-otanta hyödyttää etenkin kielioppiteorian validoimista: pätevän kielioppiteorian on kyettävä selittämään myös kaikkein harvinaisimmat kielenpiirteiden tyypit, ja ne löytyvät parhaiten variaatio-otannan avulla (Miestamo, Bakker & Arppe 2016, 233–234). Todennäköisyysotantaa puolestaan käytetään universaalien tutkimuksessa, kun halutaan testata kielenpiirteen jakaumia tilastollisesti. Tällöin tavallinen lähestymistapa otantaan on satunnaisotanta. Klassiset tilastomenetelmät vaativat, että tutkimusyksiköillä, joka

typologiassa tarkoittaa yleensä kieliä, on sama todennäköisyys tulla valituksi otokseen; tähän tavoitteeseen päästään satunnaisotannan avulla. Lisäksi tutkimusyksiköiden pitäisi olla toisistaan riippumattomia. Toisista kielistä riippuvat piirteet sotkevat tilastollisten päätelmien tekemistä, joten lähisukukielet ja lähinaapurit tulisi sulkea otoksen ulkopuolelle. Pohjimmiltaan otantamenetelmien valintaa ohjaa siis tutkimuksen tavoite.

Riippumatta otantamenetelmästä typologisen otannan pääperiaatteenä on, että otos edustaisi perusjoukkoa (esim. tällä hetkellä puhuttavia maailman kieliä) ja sen monimuotoisuutta mahdollisimman tasaisesti ja kattavasti. Tyypillisesti kattavaan ja tasaiseen edustukseen pyritään käyttämällä ositettua otantaa (engl. *stratified sampling*), jonka avulla tunnettuja vinoumia kielenpiirteiden jakaumissa pystytään rajoittamaan. Ositetussa otannassa kielet jaetaan ositteisiin, esim. kieliperheisiin ja maantieteellisiin alueisiin, joista kielet poimitaan otokseen – todennäköisyysotannassa yleensä satunnaisotannalla. Ositus voidaan tehdä joko yhdellä tai useammalla tasolla, yleensä ainakin genealogisesti eli kielisukulaisuuden perusteella. Tällöin kielet jaetaan kielikuntiin jonkin genealogisen luokittelun mukaisesti ja kielet valitaan otokseen esimerkiksi kielikunnan sisäisen rakenteen perusteella (ks. Rijkhoff & Bakker 1998) tai valitsemalla yksi kieli jokaisesta genuksesta. Genukset muodostavat maailmanlaajuisesti vertailukelpoiseksi tarkoitettun luokittelun tason, joka vastaa karkeasti indoeurooppalaisen kielikunnan päähaaroja, kuten germaanisista tai romaanisista kieliä (Dryer 1992). Toinen ositustaso on yleensä alueellinen. Tällöin kielet jaetaan samankokoisiin maantieteellisiin alueisiin, kuten Afrikka, Euraasia, Oseania, Australia, Pohjois-Amerikka ja Etelä-Amerikka, ja näiltä alueilta valitaan kieliä esimerkiksi suhteessa kielten genealogiseen monimuotoisuuteen kullakin alueella (Miestamo 2005).

Ositettu otanta on tyypillinen lähestymistapa otantaan kaikissa ihmistieteissä, mutta sillä on myös ongelmansa. Typologiassa keskeinen ongelma liittyy genealogiseen ositukseen: ei ole itsestään selvää, mitä genealogista luokittelua tulisi käyttää, ja otannan tulos voi riippua käytetystä luokittelusta. Toinen ositetun otannan ongelma liittyy synkroniseen ja dynaamiseen typologiaan, joita käsittelemme lyhyesti alla.

Otoksen muodostamisessa keskeinen kysymys on myös otoksen koko. Kun aineistoa analysoidaan sadoista kielistä, yhdestä kielestä ei voi analysoida kuin kapean alueen tai työmäärä kasvaa hallitsemattomaksi. Jos taas otoskoko liikkuu maksimissaan muutamassa kymmenessä kielessä, tutkimuskohteen pystyy analysoimaan syvällisemmin. Kysymyksenasettelu ja otoskoko rajaavat näin ollen omalta osaltaan typologista tutkimusta. Sopivan otoskoon löytäminen on tärkeää, koska liian pieni otos ei riitä luotettaviin yleistyksiin ja liian suuri otos taas hukkaa arvokasta tutkimusaikaa. Perkins (1989) esittää, että yleensä 100 kieltä on riittävä otoskoko useimpiin typologisiin tutkimuksiin. Asia ei kuitenkaan ole niin yksinkertainen, sillä riittävä otoskoko riippuu rakennetyyppien määrästä typologiassa: mitä enemmän rakennetyyppejä kielenpiirteen typologisessa luokittelussa on, sitä enemmän aineistoa tarvitaan. Tyyppien määrää ei liioin välttämättä tiedetä edeltä käsin, joten otoskoon minimi on parhaiten määriteltävissä silloin, kun tyyppien määrä tiedetään ennalta.

Perkinsin ohje koskien otoskokoa liittyy myös keskusteluun kielen universaaleista ja niiden todentamisesta tilastollisin menetelmin. Typologin tutkimusmateriaali tulee pääasiassa nykyään puhutuista kielistä, koska historiallisista kielimuodoista on säilynyt vain vähän dokumentteja ja ne muodostavat erittäin suppean otoksen historian saatossa puhutuista kielistä. Tämä ei itsessään ole ongelma, mutta ongelmaksi voi tulla se, millaisia oletuksia käyttämäämme otantamenetelmään liittyy. Kuten yllä todettiin, typologiassa on tapana käyttää ositettua otantaa, jotta otoskielet olisivat toisistaan riippumattomia. Tällöin otokseen valitaan tyypillisesti vain yksi kieli kustakin kieliperheestä tai genuksesta. Mutta kun valitaan yksi kieli kustakin kieliperheestä, otoksen perusteella ei voi päätellä, miten lähisukukielet muuttuvat kieliperheiden sisällä, vaan tuloksena on ainoastaan synkroninen tilannekuva siitä, miten tutkittu kielellinen rakenne on jakautunut maailman kielissä tiettyinä aikana. Jotta tutkija voi perustellusti tulkita, että tällainen synkroninen jakauma heijastaa kielellistä universaalia, hänen on oletettava, että rakennetyyppien jakaumat ovat saavuttaneet ns. tasapainojakauman (engl. *stationary distribution*). Tasapainojakauma tarkoittaa sitä, että muutokset yhdestä tyypestä toiseen tapahtuvat yhtä taajaan eri tyyppien kesken: koska kaikki muutokset

tyyppien välillä ovat yhtä yleisiä, synkroninen frekvenssijakauma ei enää muutu ajan myötä (Maslova 2000). Ongelmaksi muodostuu, ettei ole lainkaan selvää, miten tasapainojakauma voitaisiin todentaa.

Ratkaisuksi on tarjottu lähestymistapaa, jota kutsutaan *dynaamiseksi typologiaksi* (esim. Greenberg 1978; Croft 2003; Bickel 2015). Sillä on juurensa jo 1970-luvun typologisessa ajattelussa, mutta sitä kyettiin soveltamaan matemaattisesti ja tilastollisesti vasta 2000-luvulla (etenkin Maslova 2000). Dynaamisessa typologiassa luovutaan vaatimuksesta, että otoskielet olisivat toisistaan riippumattomia, eikä otantaa käytetä siihen, että suljettaisiin pois kielisukulaisuuden ja kielikontaktien sekoittava vaikutus. Tämä tarkoittaa sitä, että samasta kieliperheestä valitaan otokseen useita kieliä, jotta voidaan tutkia, millaisia diakronisia preferenssejä sukukielten kesken esiintyy. Kieliperheiden ja maantieteellisten alueiden vaikutusta voidaan sitten testata rakentamalla ne samaan tilastolliseen malliin testattavien lingvististen rakenteiden kanssa. Tällaiseen tilastolliseen tutkimukseen vaaditaan jo monimutkaisia monimuuttujamenetelmiä, kuten logistista regressiota tai yleistä monitasomallia (engl. *generalized multilevel modelling*). On syytä pitää mielessä, että dynaamisen lähestymistavan käyttö kasvattaa otoskokoa merkittävästi, koska kieliä valitaan sekä eri kieliperheistä että myös kieliperheiden sisältä. Tyypillinen otoskoko dynaamisessa typologiassa on noin 300–400 kieltä, usein enemmän (esim. Bickel 2015).

3.3. Kielimateriaalin hankinta

Kielimateriaalin hankintaan vaikuttavat oleellisesti tutkimuskysymys ja otoksen koko, ja tavallisesti materiaali hankitaankin samanaikaisesti otantaa tehdessä. Tyypillisin tutkimusmateriaali kielitypologiassa on yksittäisen kielen kielioppikuvaus tai tutkimuksen kohteena olevaa ilmiötä yksittäisessä kielessä käsittelevä tutkimus. Lisäksi sähköinen viestintä mahdollistaa vaivattoman yhteydenoton eri kielten asiantuntijoihin, jos omat lähteet eivät ole riittäviä.

Tutkimusmateriaalin saatavuudessa on kuitenkin suurta hajontaa. Suurista kielistä, kuten suomen kielestä, on olemassa useita

perusteellisia kuvauksia, kun taas pienistä ja uhanalaisista vähemmistökielistä ei välttämättä ole saatavilla muuta kuin yksittäisiä artikkeleita, jos niitäkään, ja ne riittävät vain harvoin lähteiksi omaan tutkimukseen. Eri kielikunnat ja maantieteelliset alueet poikkeavat toisistaan paljonkin siinä, miten syvällisesti ja laajasti niiden kieliä on tutkittu ja kuvattu. Etenkin Afrikassa ja Uudessa-Guineassa on paljon kieliä, joista ei ole vielä laadittu edes suppeaa kielioppikuvausta.

Tämä vinouma lähdemateriaalin saatavuudessa vaikuttaa siihen, mitä kieliä eri kielikunnista ja eri maantieteellisiltä alueilta voi ylipäättään valita otokseen. Tämän vuoksi lähdemateriaalin saatavuus on otettava huomioon otannassa alusta pitäen. Usein typologi joutuukin palaamaan otantaan ja vaihtamaan joitakin otoskieliä sen takia, että alun perin otokseen valittua kieltä ei kyetä analysoimaan, koska lähdemateriaali osoittautuu riittämättömäksi. Pahimmillaan lähteiden vaihteleva saatavuus voi johtaa siihen, että jotkin maantieteelliset alueet tulevat otoksessa aliedustetuiksi ja toiset taas vastaavasti yliedustetuiksi.

Kokeneella typologilla on paljon hiljaista tietoa siitä, mistä kielistä on ylipäättään saatavilla tutkimusmateriaalia, ja viime vuosina on kehitetty myös erittäin hyviä työkaluja helpottamaan materiaalin keräämistä. Erityisesti mainittakoon *Glottolog* (Hammarström, Forkel & Haspelmath 2019), joka on paitsi katalogi maailman kaikista kielistä, niiden sukulaisuhteista ja maantieteellisestä sijainnista myös viitetietokanta, joka sisältää noin 330 000 viitettä etenkin vähemmän tutkittujen kielten tutkimuskirjallisuuteen. Otantaa tehtäessä *Glottolog* on verraton työkalu, jonka avulla saa nopeasti käsityksen siitä, millaista tutkimuskirjallisuutta tietystä kielestä on saatavilla. Hyviä apuvälineitä otantaan ja tutkimusmateriaalin löytämiseksi ovat lisäksi *The Ethnologue* (Eberhard, Simons & Fennig 2019) ja WALS:n verkkoversio (Dryer & Haspelmath 2013). Varsinaista lähdemateriaalia on jo melko hyvin saatavilla avoimesti verkossa. Lisäksi Kansalliskirjaston finna.fi-palvelun kautta on saatavilla laadukkaita kuvauksia sadoista kielistä, mm. iso osa Mouton de Gruyterin julkaisemasta kirjasarjasta *Mouton Grammar Library*.

Typologit tekevät usein itsekin kenttätöitä, mutta pääasiassa typologista tutkimusta tehdään kuitenkin toisten kielentutkijoiden laatimien tutkimusten perusteella. Tämä ei kuitenkaan tarkoita sitä, että tutkija

sellaisenaan kopioisi lähdemateriaalissa olevan analyysin omaan tutkimukseensa. Typologi on toki paljolti riippuvainen lähdemateriaalin analyyseista, mutta kunnollinen lähdekritiikki on ensiarvoisen tärkeää, ja typologi muodostaa lähteiden pohjalta oman analyysinsä, mahdollisuuksien mukaan useita eri lähteitä samasta kielestä vertaillen. Tarkkuus analyysissa on tärkeää paitsi tutkimuksessa tehtävien yleistysten ja johtopäätösten kannalta myös sen tutkimuseettisenkin kysymyksen kannalta, että yksittäiset väärin analysoidut esimerkit voivat jäädä elämään kielitieteellisessä kirjallisuudessa väärinä käsityksinä jostakin yksittäisestä kielestä (tutkimuseetiikasta [ks. Johdannon luku 4 tk.](#)). Tutkijan on opittava arvioimaan lähteiden luotettavuutta, vertailtavuutta ja sitä, millaisesta teoreettisesta näkökulmasta kielen kuvaus on tehty. Lähteiden vertailtavuuteen liittyy esimerkiksi kysymys kirjoitetun ja puhutun kielen suhteesta. Kieliopit pohjautuvat tyypillisesti puhuttuihin muotoihin, mutta etenkin Euroopan kielten ja ylipäätään suurten kielten kielioppien taustalla vaikuttaa kirjoitettu yleiskieli. Tällaiset erot aineiston luonteessa voivat mahdollisesti vaikuttaa otoksesta tehtäviin typologisiin yleistyksiin (Himmelman 2000, 10–11).

Lähteiden käytettävyyteen liittyy myös kysymys metakielestä, jolla ne on kirjoitettu. Kuten luvussa 2 todettiin, referenssi-kieliopit kirjoitetaan nykyisin yleisesti peruskielioppiteorian termein ja käsittein, jolloin ne ovat paljon laajemman tutkijajoukon käytettävissä kuin jonkin spesifin teorian formaalilla metakielellä kirjoitetut kieliopit. Tästä ääriesimerkkinä voisi mainita 1960–1970-luvuilla tagmemiikan piirissä tuotetut kieliopit, jotka ovat teoriaa tuntemattomalle tutkijalle erittäin vaikeaselkoisia.³

Kielioppikuvaukset eivät suinkaan ole ainut mahdollinen lähde kielten laajassa vertailussa. Esimerkiksi saman alkuperäistekstin käännöksiä eri kielille sisältävien rinnakkaiskorpusten käyttö on yleistynyt viime vuosina typologiassa. Tyypillisimpiä käytettyjä käännöksiä ovat raamatunkäännökset, elokuvatekstitykset ja EU-tekstit.⁴ Rinnakkaistekstit avaavat uudenlaisia mahdollisuuksia vertailuun, ne mahdollistavat mm. tekstifrekvenssien vertailun, mutta niillä on omat rajoituksensa siinä, minkä tyyppisiin tutkimuskysymyksiin niitä voidaan käyttää. Lisäksi on muistettava, että rinnakkaistekstit ovat käännöstekstejä ja voivat tekstilajina poiketa ei-käännetyistä kielenkäytöstä ([ks. Käännöstiede](#)

[ja sen menetelmät tk.](#)). Kun rinnakkaiskorpuksen avulla vertailtavien kielten määrä kasvaa hyvin suureksi, jopa useisiin satoihin ja yli tuhan-teenkin kieleen (engl. *massively parallel corpora*), tarvitaan erityisiä metodeja, joita on kehitetty näin laajaan korpuspohjaiseen vertailuun (esim. Cysouw & Wälchli 2008).

Myös kyselylomakkeita käytetään typologiassa jonkin verran. Ne voivat olla hyödyllinen tapa kielimateriaalin hankintaan, etenkin jos tutkitavasta ilmiöstä on vaikea saada tietoa kielioppikuvausten tai tekstien avulla (ks. esim. Dahl 1985; Siewierska 1998; [myös Kysely tk.](#)) Kyselylomakkeet mahdollistavat systemaattisen datan keräämisen ja soveltuvat hyvin laadittuina erilaisiin tutkimuskysymyksiin joko ensisijaiseksi tai täydentäväksi datankeruumenetelmäksi. Kyselylomakkeissa on kuitenkin omat metodologiset ongelmansa: niillä kerätty data ei ole luonnollista kielenkäyttöä, varsinkaan jos kysely koostuu metakielestä käännettävistä lauseista; kuvista tai videoista koostuvissa kyselyissä ongelma on pienempi. Tällaisten elisitaatioon yleisesti liittyvien ongelmien lisäksi usein vielä lomakkeen täyttävä tutkija on ylimääräisenä välikätenä kyselylomakkeen laatineen tutkijan ja kielikonsultin välillä (ks. esim. Croft 2003, 29). Kyselylomakkeita, kuten mitä tahansa muutakin datankeruumenetelmää, käytettäessä on pidettävä kirkkaana mielessä, mihin niillä kerätty data soveltuu ja mihin ei.

Typologisen tutkimuksen lähteitä ja niiden asemaa kieliotoksessa auttaa ymmärtämään **dokulektin** käsite. Dokulektillä viitataan jonkin konkreettisen dokumentin esittämään kielimuotoon. Tällaisia dokumentteja ovat esimerkiksi referenssikielioipit, sanakirjat ja korpuksat. Esimerkiksi *Iso suomen kielioppi* (Hakulinen, Vilkuua, Korhonen, Koivisto, Heino- nen & Alho 2004) edustaa yhtä dokulektiä ja *Finnish: A Comprehensive Grammar* (Karlsson 2018) toista. Typologin kieliotos koostuu siis itse asiassa dokulekteistä, ja kieltenväliset yleistyket tehdään dokulektien pohjalta. Typologin tutkimuskysymys ei koske ensisijaisesti tutkittavaa ilmiötä yksittäisessä kielessä vaan otokseen valikoituneiden dokulektien pohjalta piirtyvää kieltenvälistä yleistystä. Tällä seikalla voidaan ajatella olevan myös tutkimuseettisiä ulottuvuuksia: typologin tehtävänä ei ole reifioida (eli tehdä olevaksi) tai lokeroida kieliä, mitä on hyvä korostaa väärinymmärrysten välttämiseksi.

Kielitypologiassa ei tyypillisesti olla suoraan tekemisissä kielenpuhujien kanssa, joten puhujien suostumukseen ja anonymiteettiin liittyvät tutkimuseettiset kysymykset koskevat typologiaa yleensä vain epäsuorasti. Toki toisten tutkijoiden keräämää aineistoa käytettäessä on kiinnitettävä huomiota aineiston käyttöehtoihin ja toimittava niiden mukaisesti. Kyselykaavakkeilla aineistoa kerättäessä nämä kysymykset ovat suuremmin osa tutkimusprosessia.

3.4. Analyysi ja luokittelu

Kun otos on laadittu ja otoskielistä on hankittu lähdemateriaali, seuraava askel on kielimateriaalin analysointi ja tarkasteltavana olevassa kielenpiirteessä havaittavan kieltenvälisen variaation luokittelu eri (rakenne)tyyppeihin. *Kielenpiirteellä* tarkoitetaan tässä siis tarkasteltavana olevaa, tyypillisesti kielen rakenteeseen liittyvää ilmiötä, kuten sanan morfologista perusrakennetta, transitiivisen lauseen hallitsevaa sanajärjestystä tai vaikkapa negaation ilmaisua. Typologit pyrkivät sekä kuvaamaan että selittämään kielenpiirteissä ilmenevää variaatiota, ja luokittelu on keskeinen vaihe tässä prosessissa. Esimerkiksi morfologisen perusrakenteen eri tyyppejä ovat luvussa 2 mainitut agglutinoivat, fusio-naaliset ja isoivat rakenteet, kun taas transitiivisen lauseen hallitsevan sanajärjestyksen tavallisia tyyppejä ovat ne eri järjestykset, joihin lauseen osallistujat ('subjekti' ja 'objekti') sekä verbi voidaan asettaa, kuten SVO, SOV tai VOS. Analyysi ja luokittelu kulkevat rinta rinnan, ja luokittelu tarkentuu sitä mukaa, kun uusia otoskieliä analysoidaan. Tällöin tyyppien määrittely voi päivittyä ja myös uusia tyyppejä voidaan tarvittaessa lisätä.

Luokittelusta ei koskaan tule aukoton, sillä jokin uusi otoskieli voi poiketa muista niin paljon, ettei sitä voida luokitella samaan rakennetyyppiin muiden kielten kanssa. Tyypit ovat kieltenvälisiä yleistyksiä, joten jonkin ilmiön löytyminen yhdestä tai kahdesta kielestä ei vielä oikeuta omaa tyyppiään. Toisaalta voi olla, että jossakin menneisyydessä puhutussa tai tulevaisuudessa puhuttavassa kielessä on esiintynyt tai esiintyy rakenne, jota ei esiinny missään otoskielessä tai edes nykyään

2)

mies 3A.3P-tarttua-PST.CMPL jaguaari

'Jaguaari nappasi miehen.' (Derbyshire 1985, 32)

Analyysivaiheessa huomio kiinnittyy etenkin kieltenvälisiin yhtäläisyyksiin ja eroihin. Näiden perusteella kielimateriaalia jäsennellään ja aineistossa esiintyvistä rakenteista tehdään typologinen luokittelu. Analyysiprosessin yksityiskohdat riippuvat pitkälti siitä, miten syvästi kukin otoskieli analysoidaan. Mikäli kielellinen ilmiö on melko

suoraviivaista analysoida, riittää, että kirjaa itselleen luokittelun määritelmät ja periaatteet, joita kielten analyysissa sovelletaan ja joita sen kuluessa mahdollisesti tarkennetaan. Tällöin voi otoskieliä analysoitaessa samanaikaisesti rakentaa tietokantaa (tietokannoista [ks. Laadullinen aineistopohjainen kielentutkimus tk.](#)), johon kunkin otoskielen ominaisuudet koodataan, eikä tekstimuotoon auki kirjoitettua analyysia tarvitse välttämättä tuottaa kuin niistä tapauksista, joiden analyysi vaatii eri vaihtoehtojen punnitsemista. Esimerkiksi Sinnemäen (2014a) noin 700 kielen analyysi objektin differentiaalisesta sijanmerkinnästä noudatteli pääosin tätä analyysitapaa: objektin sijanmerkinnän tyypit analysoitiin ja kirjattiin tietokantaan lyhyin sanallisin kuvauksin, minkä pohjalta varsinainen luokittelu sitten koodattiin tietokantaan ja perusteellisempi, auki kirjoitettu analyysi tehtiin vain niistä kielistä, joiden luokittelu oli vaativampaa. Toisaalta, jos luokiteltava kielellinen ilmiö vaatii syvällisempää analyysia, analyysin aukikirjoittaminen voi olla välttämätöntä jokaisen otoskielen kohdalla. Esimerkiksi Sinnemäen (2014b) analyysi sijanmerkinnän ja sanajärjestyksen kompleksisuudesta vaati perusteellisen 2–3 sivun tekstimuotoisen analyysin kustakin 50 otoskielestä.

Tunnetuimpia typologisia luokitteluja ovat sanajärjestystypologia ja jo luvussa 2 mainittu klassinen morfologinen typologia. Siitä, millaisia luokitteluja typologiassa on esitetty näiden lisäksi, saa hyvän yleiskuvan WALS-tietokantaa selaamalla. Jokainen WALS:n karttaluku sisältää typologisen luokittelun, jonka eri tyyppien maantieteellisiä jakaumia sitten esitellään kartalla. On hyvä todeta, että kielten rakenteiden ja ominaisuuksien luokittelu ei ole koskaan itse tarkoitus vailla muuta päämäärää, vaan se on yksi välttämätön vaihe tutkimusprosessissa. Luokittelun on oltava teoreettisesti perusteltu ja motivoitu. Typologinen tutkimus pyrkii paitsi kuvaamaan myös ymmärtämään kieltenvälistä variaatiota. Luokittelussa tulee siis pyrkiä kiinnittämään huomiota seikkoihin, jotka ovat relevantteja kieltenvälisten yhtäläisyyksien ja eroavaisuuksien selittämisessä. Esimerkiksi negaation ilmaisu voitaisiin jakaa tyypeihin sen mukaan, miten monta vokaalia negaatiota ilmaisevassa morfeemissa esiintyy. Tällainen luokittelu ei kuitenkaan auttaisi ymmärtämään negaatiota eikä tarjoaisi vastauksia mihinkään teoreettisesti kiinnostavaan negaatiota koskevaan tutkimuskysymykseen. Sen sijaan esimerkiksi

luokittelu tyyppeihin sen mukaan, miten kieltolauseet eroavat rakenteeltaan vastaavista myöntölauseista, voidaan yhdistää negaation funktionaaliin ominaisuuksiin, minkä avulla voidaan löytää motivaatioita eri rakennetyypeille (ks. Miestamo 2005).

3.5. Rakennetyyppien jakaumien tarkastelu

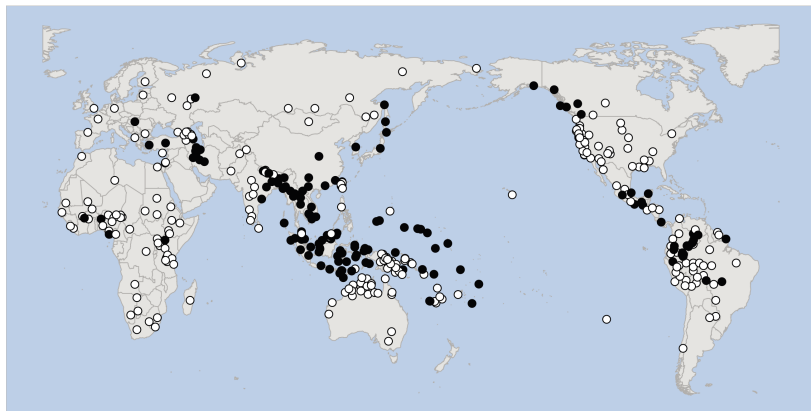
Luokittelu on ensimmäinen vaihe, kun kieltenvälisestä variaatiosta tehdään yleistyksiä. Luokittelun pohjalta voidaan sitten siirtyä tarkastelemaan tutkittavan kielenpiirteen eri rakennetyyppien yleisyyttä, niiden genealogista (eli kielikuntakohtaista) ja maantieteellistä jakautumista sekä mahdollisia korrelaatioita muiden kielenpiirteiden kanssa. Yleensä tyyppien voidaan havaita esiintyvän epätasaisesti tarkasteltaessa maailmanlaajuisia frekvenssejä tai niiden maantieteellisiä jakaumia, ja typologin tehtävänä on selvittää, mistä tämä johtuu.

Kielenpiirteiden epätasainen jakauma voi johtua genealogisista tai maantieteellisistä syistä eli siitä, että jokin tyyppi on yleisempi kuin muut tietyssä kieliperheessä tai tietyllä maantieteellisellä alueella. Näissä tapauksissa voidaan sanoa, että frekvenssien epätasainen jakauma johtuu historiallisista tapahtumista yksittäisissä kieliperheissä, joissa tyyppien yleisyys johtuu sen periytymisestä, tai yksittäisillä alueilla, joissa sama tyyppi on levinnyt kielestä toiseen kielikontaktin vaikutuksesta. Tutkimus on osoittanut, että suhteellisen harvan kielenpiirteen kohdalla eri tyypit ovat jakautuneet tasaisesti eri maanosissa, mikä kertoo kielikontaktien ja osittain myös kielisukulaisuuden vaikutuksesta tyyppien levinneisyydessä.

Kielenpiirteiden levinneisyyttä tarkastellaan typologiassa yleensä myös käyttämällä karttoja, joiden avulla tyyppien maantieteellisiä jakaumia on mahdollista helposti hahmottaa. Yllä mainittu WALS-kartasto on hyvä esimerkki toisaalta siitä, miten typologit ovat alkaneet kiinnittää yhä enemmän huomiota maantieteellisiin jakaumiin, ja toisaalta siitä, miten tärkeä työkalu kartat ovat typologille. Siinä missä maailman poliittiset kartat piirretään yleensä Atlanti-keskisesti, typologiassa kartat piirretään usein Tyynimeri-keskisesti. Tämän taustalla on ajatus siitä, että

joidenkin piirteiden maailmanlaajuiset jakaumat voivat heijastaa muinaisia väestön liikkeitä ja jopa ihmislajin leviämistä Afrikasta muihin maanosiin. On esimerkiksi todettu, että Tyynenmeren ympärillä puhuttavissa kielissä on samankaltaisuutta useissa kielenpiirteissä, joiden kohdalla muualla maailmassa esiintyy enemmän variaatiota (Nichols 1992); tästä esimerkkinä on numeraaliklassifikaattorien jakauma kuviossa 1.

Numeraaliklassifikaattorien jakauma maailman kielissä



Kuvio 1. Kartta numeraaliklassifikaattorien (Gil 2013) esiintymisestä maailman kielissä (valkoinen = ei esiinny; musta = esiintyy).⁶

Kielisukulaisuuden ja kielikontaktien lisäksi – ja myös niistä riippumatta – on mahdollista, että rakennetyyppien maailmanlaajuiset frekvenssijakaumat viittaavat kielelliseen universaaliin. Universaalit ovat tavallisimmin tilastollisia preferenssejä, siis havaintoja siitä, että tietyn kielenpiirteen jokin rakennetyyppi on muita yleisempi. Esimerkiksi lauseen sanajärjestyksen suhteen subjekti yleensä edeltää objektia, mikä tarkoittaa sitä, että rakennetyyppi ”subjekti edeltää objektia” on yleisempi kuin tyyppi ”objekti edeltää subjektiä” (esim. Bickel, Witzlack-Makarevich, Choudhary, Schlesewsky & Bornkessel-Schlesewsky 2015). Tämä on esimerkki tilastollisesta universaalista, mutta universaalit voivat olla myös absoluuttisia. Absoluuttiset universaalit koskevat kaikkia ihmiskieliä (esim. kaikilla kielillä on CV-tavuja). Teoreettisesti

kiinnostavia absoluuttisia universaaleja ei tutkimuksessa ole juurikaan löydetty, ja nekin, joiden on väitetty olevan absoluuttisia, ovat saaneet osakseen kovaa kritiikkiä (mm. Evans & Levinson 2009). Absoluuttisten universaalien löytäminen typologisista työkaluista voi olla jopa mahdotonta (Piantadosi & Gibson 2014). Funktionaalis-typologisessa tutkimuksessa tehdään tyypillisesti tilastollisia yleistyksiä kielistä ja oletetaan lingvististen rakenteiden jakautuvan probabilistisesti eli esiintyvän eri todennäköisyyksillä (ks. Bickel 2007).

Typologisiin yleistyksiin liittyy myös yleinen harhakäsitys koskien vastaesimerkkejä. Toisinaan etenkin ei-typologit voivat ajatella niiden voivan kumota typologiset yleistykset. Yksittäiset vastaesimerkit ovat potentiaalisesti ratkaiseva ongelma kuitenkin vain absoluuttisille universaaleille, sillä vastaesimerkit otetaan huomioon tilastollisessa testaamisessa. Vasta siinä tapauksessa, että uusia vastaesimerkkejä löytyy merkittävä määrä eri kielikunnista ja eri alueilta otoksen ulkopuolelta, voidaan tilastollinen typologinen yleistys kumota – aivan samoin periaattein kuin muissakin tieteissä.

3.6. Typologiset korrelaatiot

Korrelaatiot ovat kieltenvälisiä yleistyksiä kahden tai joskus useamman loogisesti toisistaan riippumattoman kielenpiirteen välisestä suhteesta. Riippumattomuus tarkoittaa tässä sitä, että kielenpiirteiden määrittelyt eivät ole keskenään päällekkäisiä vaan määrittelmien avulla rajataan tutkittavaksi eri kielenpiirteet, kuten vaikkapa sanojen keskinäinen järjestys ja sanojen taivutus. Myös korrelaatiot voivat päteä kaikkiin kieliin (= absoluuttiset universaalit) tai ne voivat päteä suureen osaan kieliä (= tilastolliset universaalit). Typologiset korrelaatiot on perinteisesti esitetty implikationaalisina universaaleina eli väitteinä, jotka kertovat kielenpiirteiden välisestä riippuvuussuhteesta implikaation muodossa $p > q$ (luetaan ”jos p , niin q ”) (Greenberg 1963; ks. myös Miestamo 2010). Tunnetuimpia tällaisia implikationaalisia universaaleja ovat sanajärjestystä koskevat yleistykset, joissa eri lause- ja lauseketyyppien sanajärjestysominaisuudet ovat riippuvuussuhteessa keskenään. On

esimerkiksi havaittu, että jos kielen hallitseva sanajärjestys on VSO, sen adpositiot ovat yleensä prepositioita, ja jos taas kielen hallitseva sanajärjestys on SOV, sen adpositiot ovat postpositioita (ks. Greenberg 1963; Dryer 1992). Tässä on itse asiassa kyse kahdesta erillisestä implikatio-naalisesta universaalista:

- 3) VSO > prepositioita
SOV > postpositioita

Eri rakenteiden sanajärjestysominaisuuksien välillä on havaittu laa-jempia korrelaatioita niin, että voidaan puhua toisaalta VO- ja toisaalta OV-kielistä, joissa verbin ja objektin järjestyksen kanssa korreloivat mm. taulukossa 1 lueteltujen konstituenttien keskinäiset järjestykset.

verbi, objekti	VO	OV
adpositio, substantiivi	AdpN (=prepositiot)	NAdp (=postpositiot)
omistettu (=N), omistaja (=G)	NG	GN
substantiivi, relatiivilause	NRel	RelN

Taulukko 1. Sanajärjestyskorrelaatioita.

Kun implikationaalisissa universaaleissa yhdistyy vähintään kaksi toisis-taan loogisesti riippumatonta kielenpiirrettä, ne ovat askel partiaalises-ta, vain yhtä kielen ilmiötä koskevasta typologiasta kohti holistisempaa typologiaa. Voidaan ajatella, että partiaalisen ja holistisen typologian välillä on jatkumo: voimme yhdistää korrelaatioiden kautta kielen-piirteitä toisiinsa ja päästä lähemmäs kielen rakenteen kokonaiskuva-a. Kielitypologian yksi keskeinen tavoite onkin selvittää, millä tavoin ja missä määrin kielenpiirteet korreloivat keskenään. Kuitenkin sellainen typologinen parametri tai parametrien joukko, joka jotenkin kokonais-valtaisesti luonnehtisi kutakin yksittäistä kieltä, näyttää nykytietämyksen valossa utopistiselta tavoitteelta. Tällaisena voidaan pitää esimerkiksi 1800-luvulla esiintynyttä ajatusta, että kielen morfologinen tyyppi ker-toisi kielestä jotakin yleisempää.

Implikationaaliset hierarkiat ovat useampien kielenpiirteiden välisiä riippuvuusketjuja, siis oikeastaan peräkkäin aseteltuja implikationaalisia universaaleja. Esimerkki implikationaalisesta hierarkiasta on lukuhierarkia, jonka esitämme tässä yksinkertaistetussa muodossa:

4) YSIKKÖ < MONIKKO < DUAALI < TRIAALI

Jos kielessä esiintyy jokin tällaisessa hierarkiassa mainittu ominaisuus, hierarkia kertoo, että silloin myös kaikki tästä vasemmalle olevat ominaisuudet esiintyvät. Ominaisuuden esiintyminen siis implikoi kaikki siitä nuolen suuntaan olevat ominaisuudet. Lukuhierarkian tapauksessa, jos esimerkiksi todetaan, että kielessä on (kahta tarkoitetta ilmaiseva) duaalikategoria, voidaan hierarkian mukaisesti päätellä, että siinä on myös monikko ja yksikkö. Hierarkian mukaan duaalin läsnäolo kielessä ei kuitenkaan kerro mitään siitä, onko kielessä mahdollisesti (kolmea tarkoitetta ilmaiseva) triaali.

Erityisen keskeiseksi kielissä vaikuttavaksi hierarkiaksi on osoittautunut elollisuushierarkia. On esitetty, että se säätelee erilaisia kieliopin ilmiöitä eri kielissä ja kiteyttää esimerkiksi objektin merkintää koskevan yleistyksen (Croft 2003; Sinnemäki 2014a). Tämän yleistyksen mukaan jos hierarkiassa alempana esiintyvä referentti merkitään objeksisijalla jossakin kielessä, myös kaikki hierarkiassa ylempänä esiintyvät referentit merkitään objeksisijalla:

5) IHMINEN < MUU ELOLLINEN < ELOTON

Hierarkia kertoo myös sen, että elolliset objektit merkitään sijalla useammin kuin elottomat objektit.

Vaikka implikationaaliset universaalit ovatkin olleet suosittu menetelmä typologisten korrelaatioiden esittämiseen, niiden käytössä on tiettyjä ongelmia (Cysouw 2005). Yksi ongelma liittyy siihen, että jotkin implikationaalisten universaalien kannalta kiinnostavat jakaumat eivät tilastotieteen näkökulmasta ole lainkaan kiinnostavia. Oletetaan esimerkiksi kaksi kuvitteellista sadan kielen jakaumaa (ks. taulukko 2), jotka esittävät piirteiden A ja B esiintymistä. Perinteisessä typologiassa

molemmat jakaumat ovat periaatteessa yhtä kiinnostavia, koska ne viittaavat piirteiden A ja B väliseen korrelaatioon: kummassakaan otoksessa ei esiinny lainkaan rakennetyyppien A+ ja B– yhdistelmää.

(a)	A+	A–	(b)	A+	A–
B+	15	70	B+	25	50
B–	0	15	B–	0	25
$p = 0.12$			$p = 0.0003$		

Taulukko 2. Implikationaaliset universaalit ja tilastollinen testaaminen. (+ = piirre esiintyy kielessä; – = piirre ei esiinny kielessä; p = Fisherin tarkalla testillä saatu todennäköisyys jakaumasta.)

Nämä jakaumat voitaisiin esittää kahden implikationaalisen universaalin avulla: $A+ > B+$ (eli jos kielessä esiintyy piirre A, siinä esiintyy myös piirre B) ja $B- > A-$ (eli jos kielessä ei esiinny piirrettä B, siinä ei esiinny myöskään piirrettä A). Kuitenkin jos näitä jakaumia testataan tilastollisesti (tässä Fisherin tarkalla testillä), ainoastaan jakauma (b) poikkeaa odotetusta jakaumasta tilastollisesti merkitsevällä tavalla ($p < 0.05$, jota käytetään yleisesti kielentutkimuksessa tilastollisen merkittävyyden rajana), kun taas jakauma (a) ei ($p > 0.05$). Näin ollen vain jakauma (b) on tilastollisessa mielessä kiinnostava. Vaikka perinteisessä typologiassa on usein käytetty tilastollisia menetelmiä, implikationaalisia universaaleja ei ole aina todennettu tilastollisesti vaan ainoastaan frekvenssien perusteella. Tämä lienee yksi syy siihen, miksi modernissa typologiassa typologisia korrelaatioita ei esitetä enää kovin usein (pelkkien) implikationaalisten universaalien vaan edistyneempien tilastollisten menetelmien avulla. Erityisesti monimuuttujamenetelmien, kuten erilaisten regressiomenetelmien, käyttö onkin lisääntynyt huomattavasti typologiassa (mm. Cysouw 2010; Sinnemäki 2011; Bickel 2015). Näiden menetelmien etuna on se, että kielisukulaisuuden ja kielikontaktien vaikutuksen suuruutta voidaan erikseen testata tilastollisesti. Esimerkiksi Sinnemäki (2011) osoitti, että kielenpiirteiden jakaumaan voivat vaikuttaa niin muut kielenpiirteet kuin myös kielisukulaisuus ja kielikontaktit, mutta kuitenkin toisistaan riippumatta.

Viime vuosina typologit ovat kiinnostuneet myös kielenpiirteiden suhteesta kielenulkoisiin tekijöihin. Kielenpiirteiden jakaumia on tutkittu esimerkiksi suhteessa puheyhteisön kokoon (Lupyan & Dale 2010; Sinnemäki & Di Garbo 2018), kielen puhuma-alueen ilmastoon (Everett, Blasi & Roberts 2015) ja geneettisiin tekijöihin (Dediu & Ladd 2007). Kielentutkimuksessa on perinteisesti ajateltu, että kielenpiirteet eivät vaihtele tällaisten tekijöiden suhteen, ja käytännössä 1900-luvun loppuun asti tutkimuksessa vältettiin tällaisia tutkimusaiheita muutamia yksittäisiä poikkeuksia lukuun ottamatta. 2000-luvun alusta lähtien myös kielenulkoiset sosiaaliset, ekologiset ja biologiset tekijät ovat herättäneet kiinnostusta ja keskustelua typologien kesken. Kun suuntaamme katseemme kielenpiirteiden ja kielenulkoisten tekijöiden suhteeseen, olemme tekemisissä kieltenvälisten yleistysten selittämisen kanssa. Seuraavassa luomme lyhyen katsauksen selittämiseen kielitypologiassa.

3.7. Tulosten selittäminen

Tutkimusprosessin viimeinen vaihe on tulosten selittäminen. Kuten yllä todettiin, typologista tutkimusta tehdään tavallisimmin funktionalistisessa viitekehyksessä. Selityksiä etsitään siis tyypillisesti kielenulkoisista tekijöistä, jotka vaikuttavat ilmiön syntyyn tai sen mukautumiseen ja sitä kautta kielten synkroniseen tilaan. Perinteiset funktionaaliset selitykset ammentavat semantiikasta, pragmatiikasta ja yleisistä kognitiivisista tekijöistä, kuten prosessointiin ja havaintoon liittyvistä seikoista. Viime aikoina on kiinnitetty myös enenevässä määrin huomiota kielen ekologiaan eli niihin lingvistisiin, sosiaalisiin, kulttuurisiin, historiallisiin, maantieteellisiin ja fyysisiin ympäristöihin, joissa kutakin kieltä käytetään ja joiden kanssa kieli on vuorovaikutuksessa (esim. Haugen 1972; Mühlhäusler 1996).

Yleisimpiä funktionaalisia selityspenusteita ovat ekonomia ja ikoniisuus (ks. erityisesti Croft 2003). Ekonomia tarkoittaa kielellisten rakenteiden ja prosessoinnin taloudellisuutta: suositaan sellaisia rakenteita, jotka päästävät puhujan mahdollisimman vähällä vaivalla. Ekonomia on ollut suosittu selityspenuste kielentutkimuksessa jo pitkään (esim.

Jespersen 1922, 261–264; Zipf 1949; Martinet 1955). Ikonisuus taas pyrkii mahdollisimman hyvään vastaavuuteen muodon ja merkityksen välillä ja toimii näin ekonomian kanssa kilpailevana motivaationa ja sen vastavoimana (esim. Haiman 1980). Esimerkki ekonomian ja ikonisuuden välisestä kilpailusta on morfologisen typologian eri tyyppien välinen kamppailu: ikonisuus pyrkii luomaan ja säilyttämään agglutinaatiivista rakennetta, jossa jokaista merkitystä vastaa yksi muoto, kun taas ekonomisuus pyrkii rikkomaan tätä ideaalia mm. lyhentämällä muotoja ja fuusioimalla ja assimiloimalla niitä toisiinsa. Ekonomia vaikuttaa etenkin kielen yleisimpien sanojen ja muotojen ominaisuuksiin, ja muotojen frekvenssiin nojaavat ekonomiapohjaiset selitykset ovat hyvä esimerkki käyttöpohjaisen lähestymistavan keskeisyydestä typologiassa (ks. Greenberg 1966; Haspelmath 2005; 2008).

Myös kielen prosessoinnilla tai omaksumisella on pyritty selittämään useita typologisia yleistyksiä (ks. Sinnemäki 2014c). Etenkin edellisessä luvussa mainittuja sanajärjestyskorrelaatioita on selitetty kielen prosessoinnilla. Hyvin yleisellä tasolla nämä selitykset perustuvat siihen, että kun samanlaisessa syntaktisessa asemassa olevat elementit sijoittuvat johdonmukaisesti samaan järjestykseen kussakin kielessä, kielen syntaktinen säännönmukaisuus lisääntyy ja kielen prosessointi helpottuu. Tunnetun tulkinnan näistä selityksperusteista on esittänyt Dryer (1992), ja vaihtoehtoisen prosessointimallin tarjoaa Hawkins (2004).

Poikkeavat rakenteet selitetään tyypillisesti historiallisilla tekijöillä. Niiden esiintyminen voidaan selittää joko kielikontakteilla tai poikkeavalla innovaatiolla satunnaisissa kielikunnissa. Esimerkiksi SVO-kielet, joissa on postpositioita, esiintyvät tyypillisesti sellaisilla maantieteellisillä alueilla, joilla eri sanajärjestystyyppejä edustavat kielet kohtaavat. Suomi on tästä hyvä esimerkki: suomalais-ugrilaisilla kielillä transitiivisen päälauseen hallitseva sanajärjestys on tavallisimmin SOV ja adpositiot ovat postpositioita. Suomessa transitiivisen päälauseen hallitseva sanajärjestys on kuitenkin muuttunut SVO:ksi, koska suomi on ollut pitkään kontaktissa germaaniin, balttilaisiin ja slaavilaisiin kieliin, joilla on SVO-sanajärjestys ja prepositioita.

4. Yhteenveto: typologia kielitieteen kentässä

Kielitypologia on maailman kielten vertailevaa tutkimusta, mutta se ei toki ole ainut kielentutkimuksen ala, jossa tehdään kieltenvälistä vertailua tai jossa pyritään sanomaan jotakin yleistä ihmiskielestä. Yksi typologian lähialoista on luonnollisesti kontrastiivinen kielentutkimus, jossa myös verrataan kieliä toisiinsa ja pohditaan muun muassa, millä edellytyksillä kielten vertailu on mahdollista. Nämä alat kuitenkin eroavat toisistaan sekä tavoitteiltaan että menetelmiltään. Siinä missä kielitypologiassa tutkitaan systemaattisesti kieltenvälisiä eroja ja samankaltaisuuksia hyvinkin erilaisissa kielissä, kontrastiivisessa kielentutkimuksessa keskitytään tyypillisesti vertailemaan kahta kieltä, jotka ovat joko toisilleen sukua tai joita puhutaan maantieteellisesti samalla alueella. Kielitypologian perimmäinen tarkoitus on ymmärtää paremmin ihmiskielen luonnetta, joten sen tavoitteet ovat teoreettiset ja liittyvät pitkälti perustutkimukseen. Kontrastiivisen kielentutkimuksen tavoitteet ovat selvemmin soveltavia, ja tällainen tutkimus palvelee usein esimerkiksi kieltenopetuksen tarpeita.

Kielitypologian lähialaksi voidaan katsoa myös (formalistinen) kielioppiteoria, sillä se on typologian tapaan kiinnostunut kielen universaaleista. Toisinaan onkin ajateltu, että kielitypologia ja chomskylainen universaalikielioppi muodostavat saman kolikon eri kääntöpuolet. Näiden alojen lähtökohdat ovat kuitenkin hyvin erilaiset, vaikka niiden kiinnostuksen kohteet osittain risteävätkin. Tässä artikkelissa olemme tarkastelleet etenkin funktionaalis-typologista lähestymistapaa kielentutkimukseen, ja tämän alan teoreettiset lähtökohdat ovat kielen käytössä ja kielen funktiossa maailman eri kielissä. Etenkin generatiivisessa kielitieteessä lähtökohta on kuitenkin kielen oletetussa biologisessa perustassa ja universaalissa kielikyvyyssä, ja sen parissa on ollut tapana tutkia yhtä kieltä (usein englantia) syvällisesti toisin kuin kielitypologiassa, jossa on tapana tutkia kymmeniä tai jopa satoja kieliä samanaikaisesti. Bickel (2007) on jopa esittänyt, että funktionaalis-typologinen kielentutkimus on osa ihmiskulttuurin tutkimusta (laajasti ymmärrettynä), kun taas universaalikielioppi on osa kognitiotiedettä. Todettakoon

vielä, että kielen universaaleja ominaisuuksia tutkitaan myös monien kognitiivis-funktionalistisesti suuntautuneiden kielioppiteorioiden parissa (esim. langackerilainen kognitiivinen kielioppi ja goldbergiläinen konstruktiokielioppi; [ks. Kieliopin tutkimus ja kielioppiteoriat tk.](#)), ja nämä lähestymistavat ovat teoreettisesti hyvinkin yhteensopivia typologis-funktionalistisen ajattelun kanssa.

Kielitypologialle läheisimmät tutkimusalat löytyvät kuitenkin ns. diversiteettilingvistiikan piiristä. Tällä termillä on viime aikoina alettu kutsua toisiinsa läheisessä suhteessa olevia kielitieteen aloja, joiden yhteisenä tavoitteena on maailman kielellisen diversiteetin kartoittaminen ja ymmärtäminen. Näitä aloja ovat kielitypologian lisäksi vähemmän tutkittujen kielten dokumentointi ja kuvaus, antropologinen lingvistiikka, historiallis-vertaileva kielentutkimus, areaalilingvistiikka ja kielikontaktien tutkimus (ks. ko. lukuja tk.). Erityisesti kielitypologia ja deskriptiivinen kielentutkimus ovat läheisessä suhteessa. Kuten jo johdannossamme totesimme, niiden välillä vallitsee riippuvuussuhde: siinä missä deskriptiivinen kielentutkimus tuottaa tietoa typologeille aiemmin heikosti tunnetuista kielistä ja niiden rakenteista, kielitypologinen tieto auttaa kenttalingvistiä suhteuttamaan käsillä olevan kielen rakenteen analyysin siihen, mitä maailman kielistä tiedetään, ja näin ymmärtämään tutkimansa kielen ilmiötä paremmin. Näidenkin alojen lähtökohdat ja menetelmät eroavat kuitenkin toisistaan. Deskriptiivinen kielentutkimus pyrkii ymmärtämään ja kuvaamaan yksittäisen kielen kielioppia, ja tyypillisesti kunkin kielet kategoriat on kuvattava niiden omilla ehdoilla (ks. luku 2). Kielitypologiassa typologi pyrkii luomaan käsitteitä, joiden avulla kieltenvälinen vertailu olisi mahdollista huolimatta yksittäisten kielten kategorioiden erilaisuudesta. Kielitypologi ei siis ensisijaisesti pyri sanomaan mitään jonkin yksittäisen kielen ominaisuuksista, vaan hänen tutkimuskysymyksensä koskevat kieltenvälistä vaihtelua – kielten erilaisuutta ja samuutta ja niiden taustalla vaikuttavia tekijöitä.

Tässä artikkelissa olemme esitelleet kielitypologiaa yhtenä lähestymistapana kielentutkimukseen. Esittelimme kielitypologian keskeisiä teoreettisia taustoja, sen tutkimushistorian pääpiirteet sekä typologisen tutkimusprosessin eri vaiheet. Lisäksi olemme käsitelleet typologian

suhdetta muihin kielentutkimuksen lähialoihin. Kielitypologia kehittää menetelmiä maailmanlaajuiseen kieltenvälisen variaation tutkimukseen. Sen perimmäiset tavoitteet liittyvät ihmiskielen luonteen selvittämiseen, ja se on tärkeä lähestymistapa yleisen kielitieteen keskeisiin kysymyksiin. Typologinen tutkimus luo uutta tietoa kielten eroista ja samankaltaisuuksista ja niihin vaikuttavista tekijöistä, joten se tarjoaa tärkeää tietoa selvitetessä, miten kieli varioi, miten se muuttuu ja miten se on syntynyt.

Aiheesta lisää:

Aikhenvald, Alexandra & Dixon, R. M. W. (toim.). 2017. *The Cambridge Handbook of Linguistic Typology*. Cambridge: Cambridge University Press.

Comrie, Bernard. 1989. *Language Universals and Linguistic Typology*. (2. p.). Chicago (IL): The University of Chicago Press.

Croft, William. 2003. *Typology and Universals*. (2. p.). Cambridge: Cambridge University Press.

Dryer, Matthew S. & Haspelmath, Martin (toim.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [verkkoaineisto]. [viitattu 25.4.2018]. Saatavissa: <https://wals.info>.

Shopen, Timothy (toim.). 2007. *Language Typology and Syntactic Description*. (3 nidettä, 2. p.). Cambridge: Cambridge University Press.

Song, Jae Jung (toim.). 2010. *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press.

Velupillai, Viveka. 2012. *An Introduction to Linguistic Typology*. Amsterdam: John Benjamins.

VIITTEET

- 1 Kiitämme artikkelimme toimittajia ja arvioijia hyödyllisistä kommenteista. Kaius Sinnemäki on saanut rahoitusta artikkelin kirjoittamiseen Suomen Akatemialta (hanke nro 296212) sekä Euroopan tutkimusneuvostolta (ERC) Euroopan unionin Horisontti 2020-tutkimus- ja innovaatio-ohjelmasta (sopimus nro 805371).
- 2 "languages could differ from each other without limit and in unpredictable ways".
- 3 *Tagmeemin* käsite syntyi strukturalismin parissa (Bloomfield 1933), mutta sen teki tunnetuksi Kenneth Pike (1912–2000), joka kehitti tagmeemianalyysistä kattavamman kielen kuvauksen teorian (esim. Pike 1982). Lähestymistavan vaikeaselkoisuus johtuu osittain siitä, että siinä korostettiin kategorioiden kielikohtaisuutta siihen pisteeseen asti, että myös kategorioita kuvaaville termeille annettiin kielikohtaiset, esimerkiksi numeroista ja kirjaimista koostuvat nimet. Lisäksi kuvauksen taulukoiden ja formalismien tukena käytettiin usein vain hyvin niukalti suorasanaisia selityksiä.
- 4 Katso lisää Jörg Tiedemannin ylläpitämästä OPUS-verkkosivustosta (<http://opus.lingfil.uu.se/>), joka listaa avoimesti saatavilla olevia rinnakkaiskorpuksia.
- 5 Glossaussäännöt löytyvät osoitteesta <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>. Esimerkissä käytetyt lyhenteet ovat: A = agentti, lauseen tekijä; P = patientti, teon kohde; C_{MPL} = päättynyt aspekti; PST = menneen ajan tempus.
- 6 Kuva on tehty R-ohjelmalla (R Core Team 2018) käyttämällä sen paketteja ggplot2 (Wickham 2016), ggmap (Kahle & Wickham 2013) ja mapdata (Becker, Wilks & Brownrigg 2018).

KIRJALLISUUS

- Becker, Richard A., Wilks, Allan R. & Brownrigg, Ray. 2018. mapdata: Extra Map Databases. R package version 2.3.0. Saatavissa: <https://CRAN.R-project.org/package=mapdata>.
- Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11:1, 239–251.
- Bickel, Balthasar. 2015. Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. Julkaisussa: Heine, Bernd & Narrog, Heike (toim.) *The Oxford Handbook of Linguistic Analysis*. (2. p.). Oxford: Oxford University Press, 901–923.
- Bickel Balthasar, Witzlack-Makarevich, Alena, Choudhary, Kamal K., Schlesewsky, Matthias & Bornkessel-Schlesewsky, Ina. 2015. The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLoS ONE* 10:8, e0132819.
- Bloomfield, Leonard. 1933. *Language*. New York (NY): Henry Holt.
- Croft, William. 2003. *Typology and Universals*. (2. p.). Cambridge: Cambridge University Press.
- Cysouw, Michael. 2005. Quantitative methods in typology. Julkaisussa: Altmann, Gabriel, Köhler, Reinhard & Piotrowski, Rajmund G. (toim.) *Quantitative Linguistics: An International Handbook*. Berlin: Mouton de Gruyter, 554–578.
- Cysouw, Michael. 2010. Dealing with diversity: Towards an explanation of NP word order frequencies. *Linguistic Typology* 14:2, 253–287.

- Cysouw, Michael & Wälchli, Bernhard (toim.). 2008. Parallel texts. Using translational equivalents in linguistic typology. *Language Typology and Universals (STUF)* 60:2, 95–181.
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.
- Dediu, Dan & Ladd, D. Robert. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences* 104:26, 10944–10949.
- Derbyshire, Desmond C. 1979. *Hixkaryana*. Amsterdam: North-Holland.
- Derbyshire, Desmond C. 1985. *Hixkaryana and Linguistic Typology*. Dallas (TX): Summer Institute of Linguistics.
- Dixon, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. 2009–2012. *Basic Linguistic Theory*. (3 nidettä). Cambridge: Cambridge University Press.
- Dryer, Matthew. 1992. The Greenbergian word order correlations. *Language* 68:1, 81–138.
- Dryer, Matthew. 1997. Are grammatical relations universal? Julkaisussa: Bybee, Joan Haiman, John & Thompson, Sandra (toim.) *Essays on Language Function and Language Type: Dedicated to T. Givón*. Amsterdam: John Benjamins, 115–143.
- Dryer, Matthew. 2006. Descriptive theories, explanatory theories and Basic Linguistic Theory. Julkaisussa: Ameka, Felix, Dench, Alan & Evans, Nicholas (toim.) *Catching Language: The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter, 207–234.
- Dryer, Matthew S. & Haspelmath, Martin (toim.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [verkkoaineisto]. [viitattu 25.4.2018]. Saatavissa: <https://wals.info>.
- Eberhard, David M., Simons, Gary F. & Fennig, Charles D. (toim.). 2019. *The Ethnologue: Languages of the World*. (22. p.). Dallas (TX): SIL International. [verkkoaineisto]. [viitattu 16.12.2019]. Saatavissa: <https://www.ethnologue.com>.
- Evans, Nick & Levinson, Stephen. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32:5, 429–448.
- Everett, Caleb, Blasi, Damián E. & Roberts, Seán G. 2015. Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences* 112:5, 1322–1327.
- Everett, Daniel. 2004. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46:4, 621–646.
- Gil, David. 2013. Numeral classifiers. Julkaisussa: Dryer, Matthew & Haspelmath, Martin (toim.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [verkkoaineisto]. [viitattu 25.4.2018]. Saatavissa: <https://wals.info/chapter/55>.
- Graffi, Giorgi. 2010. The pioneers of linguistic typology: from Gabelentz to Greenberg. Julkaisussa: Song, Jae Jung (toim.) *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press, 25–42.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. Julkaisussa: Greenberg, Joseph H. (toim.) *Universals of Language*. Cambridge (MA): MIT Press, 73–113.

- Greenberg, Joseph H. 1966. *Language Universals, with Special Reference to Feature Hierarchies*. The Hague: Mouton.
- Greenberg, Joseph H. 1978. Diachrony, synchrony and language universals. Julkaisussa: Greenberg, Joseph H., Ferguson, Charles A. & Moravcsik, Edith A. (toim.) *Universals of Human Language I: Method and Theory*. Stanford (CA): Stanford University Press, 61–92.
- Haiman, John. 1980. The iconicity of grammar: Isomorphism and motivation. *Language* 56:3, 515–540.
- Hakulinen, Auli, Vilkuna, Maria, Korhonen, Riitta, Koivisto, Vesa, Heinonen, Tarja Riitta & Alho, Irja. 2004. *Iso suomen kieliooppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Hammarström, Harald, Forkel, Robert & Haspelmath, Martin. 2019. *Glottolog 4.1*. Jena: Max Planck Institute for the Science of Human History. [verkkoaineisto]. [viitattu 1.2.2019]. Saatavissa: <https://glottolog.org>.
- Haspelmath, Martin. 2005. Preface to the reprinted edition. Julkaisussa: Greenberg, Joseph H. *Language Universals, with Special Reference to Feature Hierarchies*. (Uudelleenjulkaistu p.). The Hague: Mouton, vii–xvii.
- Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11:1, 119–132.
- Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19:1, 1–33.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86:3, 663–687.
- Haspelmath, Martin, Dryer, Matthew, Gil, David & Comrie, Bernard (toim.). 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Haugen, Einar. 1972. *Ecology of Language*. Stanford (CA): Stanford University Press.
- Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Himmelman, Nikolaus P. 2000. Towards a typology of typologies. *Language Typology and Universals (STUF)* 53:1, 5–12.
- Itkonen, Esa. 1991. *Universal History of Linguistics: India, China, Arabia, Europe*. Amsterdam: John Benjamins.
- Itkonen, Esa. 1997. *Maailman kielten erilaisuus ja samuus*. Helsinki: Gaudeamus.
- Jespersen, Otto. 1922. *Language: Its Nature, Development and Origin*. London: Allen & Unwin.
- Joos, Martin. 1957. *Readings in Linguistics: The Development of Descriptive Linguistics in America since 1925*. Washington: American Council of Learned Societies.
- Kahle, David & Wickham, Hadley. 2013. ggmap: Spatial Visualization with ggplot2. *The R Journal* 5:1, 144–161.
- Karlsson, Fred. 2018. *Finnish: A Comprehensive Grammar*. London: Routledge.
- Kittilä, Seppo. 2002. *Transitivity: Towards a Comprehensive Typology*. Turku: Turun yliopisto.
- Lazard, Gilbert. 1992. Y a-t-il des catégories interlangagières? Julkaisussa: Anschütz, Susanne & Heger, Klaus (toim.) *Texte, Sätze, Wörter und Moneme: Festschrift für Klaus Heger zum 65. Geburtstag*. Heidelberg: Heidelberg Orient-Verlag, 427–434.
- Lupyan, Gary & Dale, Rick. 2010. Language structure is partly determined by social structure. *PLoS One* 5:1, e8559.

- Martinet, André. 1955. *Économie des Changements Phonétiques: Traité de Phonologie Diachronique*. Berne: Francke.
- Marttila, Annu. 2010. *A Cross-linguistic Study of Lexical Iconicity and Its Manifestation in Bird Names*. Väitöskirja. Helsinki: Helsingin yliopisto.
- Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4:3, 307–333.
- Miestamo, Matti. 2000. Towards a typology of standard negation. *Nordic Journal of Linguistics* 23:1, 65–88.
- Miestamo, Matti. 2003. *Clausal Negation: A Typological Study*. Väitöskirja. Helsinki: Helsingin yliopisto.
- Miestamo, Matti. 2005. *Standard Negation: The Negation of Declarative Verbal Main Clauses in a Typological Perspective*. Berlin: Mouton de Gruyter.
- Miestamo, Matti. 2010. Joseph Greenberg ja kielitypologia. Julkaisussa: Haddington, Pentti & Sivonen, Jari (toim.) *Kielentutkimuksen modernit klassikot*. Helsinki: Gaudeamus, 153–182.
- Miestamo, Matti. 2013. Kielten vertailu kielitypologisessa tutkimuksessa. Julkaisussa: Kolehmainen, Leena, Miestamo, Matti & Nordlund, Taru (toim.) *Kielten vertailun metodiikka*. Helsinki: Suomalaisen Kirjallisuuden Seura, 27–55.
- Miestamo, Matti, Bakker, Dik & Arppe, Antti. 2016. Sampling for variety. *Linguistic Typology* 20:2, 233–296.
- Mühlhäusler, Peter. 1996. *Linguistic Ecology: Language Change and Linguistic Imperialism in the Pacific Region*. London: Routledge.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago (IL): The University of Chicago Press.
- Pajunen, Anneli (toim.). 1998. *Kieliopillistumisesta, analogiasta ja typologiasta*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Pajunen, Anneli (toim.). 2000. *Näkökulmia kielitypologiaan*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Perkins, Revere D. 1989. Statistical techniques for determining language sample size. *Studies in Language* 13:2, 293–315.
- Piantadosi, Steven T. & Gibson, Edward. 2014. Quantitative standards for absolute linguistic universals. *Cognitive Science* 38, 736–756.
- Pike, Kenneth L. 1982. *Linguistic Concepts: An Introduction to Tagmemics*. Lincoln (NB): University of Nebraska Press.
- Ramat, Paolo. 2010. The (early) history of linguistic typology. Julkaisussa: Song, Jae Jung (toim.) *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press, 9–24.
- R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Saatavissa: <https://www.R-project.org/>.
- Rijkhoff, Jan & Bakker, Dik. 1998. Language sampling. *Linguistic Typology* 2:3, 263–314.
- Salminen, Tapani. 1996. *Typological and Historical Studies in Language by Mikko Korhonen: A Memorial Volume Published on the 60th Anniversary of his Birth*. Helsinki: Suomalais-Ugrilainen Seura.
- Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. Orlando: Harcourt Brace & co.

- Shagal, Ksenia. 2019. *Participles: A Typological Study*. Berlin: De Gruyter Mouton.
- Siewierska, Anna (toim.). 1998. *Constituent Order in the Languages of Europe*. Berlin: Mouton de Gruyter.
- Sinnemäki, Kaius. 2008. Complexity trade-offs in core argument marking. Julkaisussa: Miestamo, Matti, Sinnemäki, Kaius & Karlsson, Fred (toim.) *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins, 67–88.
- Sinnemäki, Kaius. 2011. *Language Universals and Linguistic Complexity: Three Case Studies in Core Argument Marking*. Helsinki: Helsingin yliopisto. Saatavissa: <http://urn.fi/URN:ISBN:978-952-10-7259-8>.
- Sinnemäki, Kaius. 2014a. A typological perspective on differential object marking. *Linguistics* 52:2, 281–313.
- Sinnemäki, Kaius. 2014b. Complexity trade-offs: A case study. Julkaisussa: Newmeyer, Frederick J. & Preston, Laurel B. (toim.) *Measuring Grammatical Complexity*. Oxford: Oxford University Press, 179–201.
- Sinnemäki, Kaius. 2014c. Cognitive processing, language typology, and variation. *WIREs Cognitive Science* 5:4, 477–487.
- Sinnemäki, Kaius & Di Garbo, Francesca. 2018. Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology* 9:1141. Saatavissa: <https://doi.org/10.3389/fpsyg.2018.01141>.
- Stassen, Leon. 1985. *Comparison and Universal Grammar*. Oxford: Basil Blackwell.
- Stassen, Leon. 2010. The problem of cross-linguistic identification. Julkaisussa: Song, Jae Jung (toim.) *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press, 90–99.
- Suihkonen, Pirkko. 1990. *Korpustutkimus kielitypologiassa sovellettuna udmurttiin*. Helsinki: Suomalais-Ugrilainen Seura.
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York (NY): Springer-Verlag.
- Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort*. Oxford: Addison-Wesley.